

NASA Technical Memorandum 4048

The SAMEX Vector Magnetograph

*A Design Study for a Space-Based
Solar Vector Magnetograph*

M. J. Hagyard, G. A. Gary,
and E. A. West

JUNE 1988

(NASA-TM-4048) THE SAMEX VECTOR
MAGNETOGRAPH: A DESIGN STUDY FOR A
SPACE-BASED SOLAR VECTOR MAGNETOGRAPH
(NASA) 437 P

N88-25424

CSCI 03B

H1/92 0147156

Unclas

NASA

NASA Technical Memorandum 4048

The SAMEX Vector Magnetograph

*A Design Study for a Space-Based
Solar Vector Magnetograph*

**M. J. Hagyard, G. A. Gary,
and E. A. West**
*George C. Marshall Space Flight Center
Marshall Space Flight Center, Alabama*



**National Aeronautics
and Space Administration**

**Scientific and Technical
Information Division**

1988

ACKNOWLEDGEMENTS

The writing of this report has benefited significantly from numerous informal discussions with our colleagues in the Solar Science Branch of the NASA/MSFC Space Science Laboratory. In the early phase of the study, these colleagues included E. Hildner and D. Rabin. Special thanks go to Ron Moore who wrote some of the sections on scientific rationale and to John Davis who provided constructive comments and advice relevant to the section on CCD detectors. John also was very helpful in some of the management aspects of this study.

We thank members of the Lockheed Research Laboratories for the written material on birefringent filters found in Appendix A and in the body of the text. In particular we are indebted to Bill Rosenberg, Alan Title, and Ted Tarbell for their inputs to this study and their many helpful discussions with us.

The material on solid-etalon Fabry-Perot filters was supplied by Dave Rust, T. Appourchaux, and T. Harris of the Applied Physics Laboratory at the Johns Hopkins University. We are particularly grateful to Dave Rust for his interest in this study and for his many useful contributions.

The optical design of the system and instrumental polarization analysis were accomplished with the assistance of the Center for Applied Optics at the University of Alabama in Huntsville. We thank the director of the Center, John Caulfield, and Barry Johnson, Russell Chipman, Jim McGuire, and Jim Trolinger for the support and work provided to carry out this study.

This study was supported by the Air Force Geophysics Laboratory through its Solar Research Division. We thank Don Neidig of that group for his support and for the many fruitful discussions concerning the SAMEX program.

Much of the information and material reported in this document is derived from the experience we have gained over the past decade from our research program with the MSFC solar vector magnetograph system. This research program is supported by the NASA Office of Solar and Heliospheric Physics, and we thank the chief of that Office, J. D. Bohlin, for his continuing interest and support.

The vector magnetograph program at MSFC has benefited significantly from the MSFC Center Director's Discretionary Fund Program, and we are indebted to the Marshall Center for its continuing support of our research and development program.

TABLE OF CONTENTS

	Page
I. INTRODUCTION	1
1. The SAMEX Mission	1
2. A Space-Based Solar Vector Magnetograph	5
3. Scientific Basis and Objectives	8
II. THE VECTOR MAGNETOGRAPH INSTRUMENT	48
1. Introduction	48
2. The Optical System	64
3. Analysis of Instrumental Polarization of Imaging Optics	89
4. The Polarimeter	115
5. Tunable Spectral Filter	151
6. Detector System	191
7. Overall System Performance	236
III. SPACECRAFT REQUIREMENTS	238
1. OnBoard Computer System	238
2. H-Alpha System and Image Motion Compensation ...	245
3. Experiment Observing Modes	251
4. Spacecraft Options	255
5. Telemetry	258
IV. MISSION REQUIREMENTS	263
1. Introduction	263
2. Orbital Considerations	263
3. Duration of Mission	268
4. Ground Operations and Data Management	269
V. DATA REDUCTION AND ANALYSIS	277
1. Introduction	277
2. Software for Data Analysis	277
3. Theoretical Modeling	280
VI. CONCLUSIONS AND RECOMMENDATIONS	294
1. Science Summary	294
2. Instrument Summary	297
3. Recommended Studies	298
4. Recommended Test Programs	299
5. Recommended Developments in Technology	300
6. Conclusion	300

TABLE OF CONTENTS (Cont'd)

	Page
REFERENCES	301
APPENDIX A. Parameters of Glan-Thompson Polarizers	
APPENDIX B. Narrow Bandpass Birefringent Filters	
APPENDIX C. Use of a Fabry-Perot Filter in a Space Flight Solar Vector Magnetograph	
APPENDIX D. Polarization Aberration Theory	
APPENDIX E. Interpretation of Data from the SAMEX Vector Magnetograph	
APPENDIX F. Measurement of Doppler Velocities with the SAMEX Magnetograph	

I. INTRODUCTION

This report presents the results of a pre-phase A study performed by the Marshall Space Flight Center (MSFC) for the Air Force Geophysics Laboratory (AFGL) to develop a design concept for a space-based solar vector magnetograph and hydrogen-alpha telescope. These instruments are two of the core instruments for a proposed Air Force mission, the Solar Activity Measurements Experiments (SAMEX). This mission is designed to study the processes which give rise to solar activity in the solar atmosphere and to develop techniques for predicting solar activity and its effect on the terrestrial environment.

1. The SAMEX Mission

The operations of military and civilian satellites, and systems for communications, tracking, and surveillance can be interrupted, degraded, or even endangered as a result of powerful explosions on the surface of the Sun that we know as solar flares. These spectacular eruptions release shock waves, hot plasma clouds, highly accelerated atomic nuclei, and bursts of x-rays, ultraviolet, and visible-band electromagnetic radiation into interplanetary space. When the path of propagation of these high-energy emissions intersects the Earth's magnetosphere, our terrestrial environment is impacted in various ways that may produce deleterious effects on military and civilian systems, both on the ground and in space (Figure 1). Effects from other forms of solar activity can be equally harmful to these systems. The appearance of high-speed solar wind streams from coronal holes causes disturbances that propagate through the interplanetary medium to Earth's environment. Another form of solar activity is the sudden eruption of filaments from the solar atmosphere causing interplanetary disturbances which can affect the Earth's environment in the form of geomagnetic storms.

The SAMEX mission has been developed toward advancing our scientific understanding of the physical mechanisms driving these phenomena and thereby improving current techniques for predicting these disruptive solar and geomagnetic events. Since our current understanding of solar activity holds that the solar magnetic field plays the fundamental role, any study aimed at understanding the various phenomena associated with solar activity must focus on this magnetic field. Thus the primary scientific objectives of the SAMEX mission are:

- o to place on a quantitative basis our understanding of the buildup and storage of energy in the solar magnetic field and the relationship of this field to structures at all levels in the solar atmosphere;
- o to identify the conditions which are necessary for the release of the magnetic energy in various solar transient phenomena;
- o to determine whether this knowledge is sufficient to predict where, when, how much, and in what form the magnetic energy will be released.

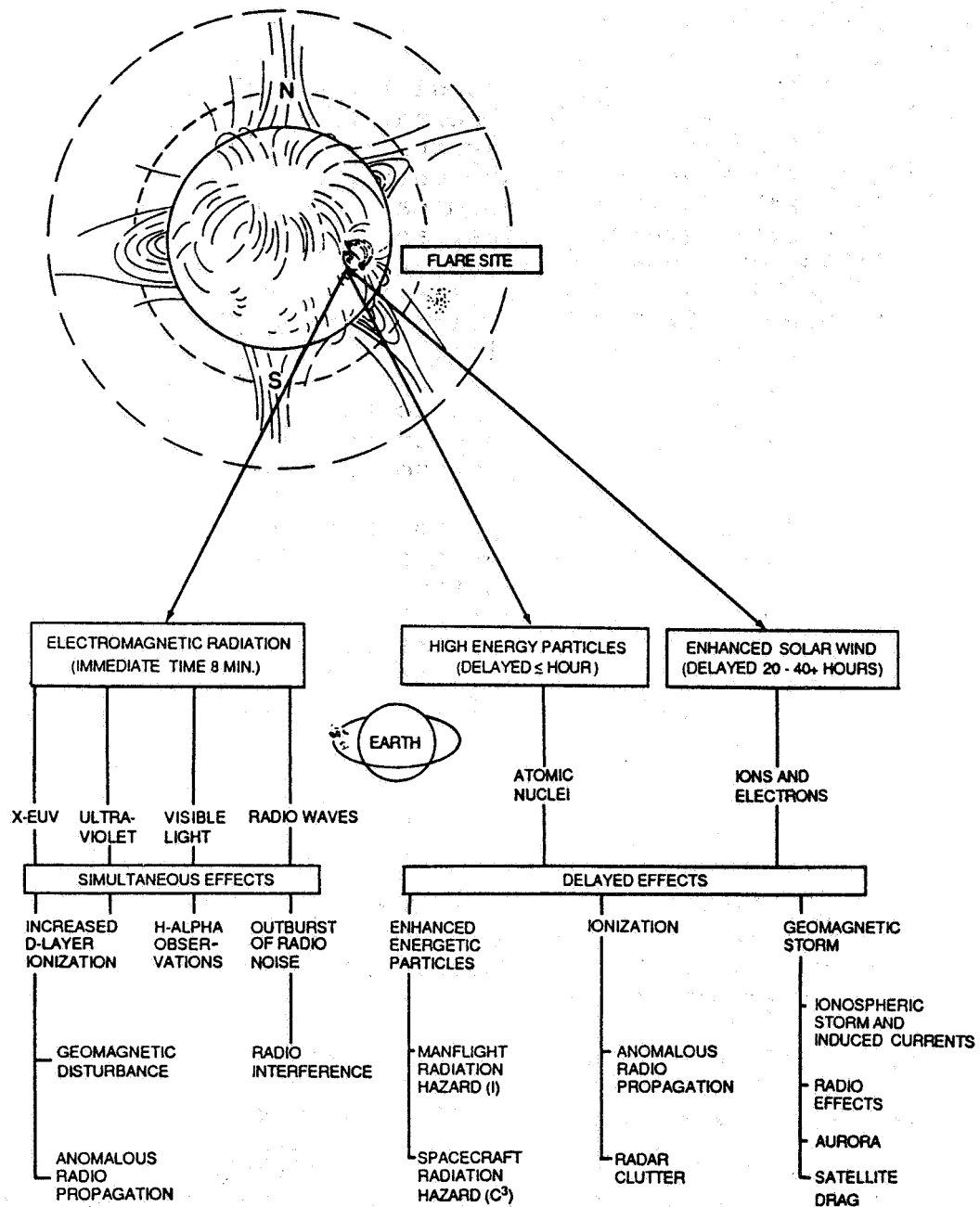


Figure 1. Solar-Terrestrial Relationships (after Ellison, 1955). The three physical mechanisms of electromagnetic radiation, high-energy particles, and enhanced solar wind resulting from a solar flare affect the near-Earth and terrestrial environment. The specific composition of, and effects from, the three physical mechanisms are shown. The electromagnetic radiation arrives almost immediately and affects radio communications. The high-energy particles and particles in the enhanced solar wind travel more slowly but produce radiation hazards and ionospheric disturbances. A vector magnetograph is designed to study the root cause of these effects: changes in the solar magnetic field at the site of the flare.

The SAMEX mission will consist of instruments to make coordinated, high-resolution observations from space of the magnetic and dynamical characteristics of solar active-region development, flares and preflare energy buildup, as well as the relationship between these phenomena and high temperature coronal plasma structures, instabilities, and convective processes in the photosphere and sub-photospheric layers. To achieve these objectives, a basic instrument package is envisaged for the SAMEX mission consisting of three imaging instruments, each operating at 0.5 arc sec spatial resolution:

- o a vector magnetograph to measure the longitudinal and transverse components of the surface magnetic field and the line-of-sight velocities of the photospheric gases;
- o an H-alpha telescope to observe the structure and dynamics of the plasma in the chromosphere that delineates the magnetic field there;
- o an x-ray/XUV telescope to define the magnetic configuration in the transition region and corona, and to measure parameters of the low density, high temperature plasma structures there.

These instruments have been chosen on the basis of present-day observations and theory, and they are generally accepted as being necessary to advance our understanding of the processes involved in the development of active regions, energy buildup, and the triggering and release of energy in flares (Neidig, 1986). The key to the problem of solar activity lies in the active-region magnetic field, which permeates all layers of the solar atmosphere. This complement of instruments for SAMEX provides a "look" at this magnetic field at all levels: the field itself in the photosphere and chromosphere, and x-ray/XUV emissions in the million-degree plasmas that fill the field structures in the transition region and corona. The SAMEX instruments will observe all of these simultaneously.

The core instruments have evolved to a sufficiently sophisticated level that development of space-based counterparts seems justified and feasible. A ground-based vector magnetograph (Hagyard et al., 1982, 1985a) has been in operation at MSFC since 1976, and x-ray/XUV telescopes have been developed and flown in space (but these telescopes did not have the spatial resolution of 0.5" that is the goal of SAMEX). The combination of these instruments into a single mission will provide the coordinated observations that we expect will revolutionize our understanding of flares and other transient solar phenomena and will provide the scientific foundations for forecasting these manifestations of solar activity that produce effects upon the near-Earth environment. SAMEX will, therefore, be of practical importance to NASA, the Department of Defense (DOD), and the National Oceanic and Atmospheric Administration (NOAA). In addition, the SAMEX science goals are central both to a broad cross section of

solar and heliospheric physicists and to a wide community of astronomers and plasma physicists whose interests encompass the Sun as but one example of the physical systems they study. SAMEX will provide the most comprehensive data yet available on solar magnetized plasmas, and by doing so it will present an opportunity for major advances in research areas that have been repeatedly cited as being fundamentally important to the future development of astrophysics. Studies of solar magnetic fields and field reconnection, plasmas, the interaction between convective processes and magnetic fields, and, specifically, flare mechanisms have been cited as major research objectives by various national committees in the following reports of the National Academy of Sciences: A Strategy for Space Astronomy and Astrophysics for the 1980's (1979), Solar-Terrestrial Research for the 1980's (1981), Astronomy and Astrophysics for the 1980's (1982), National Solar-Terrestrial Research Program (1984), The Physics of the Sun (1985).

SAMEX will constitute the first long-term, high-resolution mission coordinating observations at all levels in the solar atmosphere. It therefore answers most of the projected needs for high-resolution solar data cited by these national study groups as being vital to future development of the solar-terrestrial sciences. Moreover, SAMEX would provide simultaneous observations of the solar plasma over a wide temperature range, 6000 to greater than 2×10^6 K--a capability specifically recommended by the Field Report (Astronomy and Astrophysics for the 1980's, National Academy Press, 1982). Within NASA, NOAA, and DOD, the vastly improved understanding of solar activity processes will bring about new capabilities to predict conditions in the aerospace environment. This represents a practical application, the importance of which will increase dramatically as these agencies become increasingly reliant upon activities in space. We do not expect the SAMEX mission to provide predictive capabilities in an operational sense; rather, SAMEX will develop the prerequisite science and provide proof-of-concept for future missions by DOD and NOAA.

Within universities and national centers, as well as within NASA Centers, scientists will have immediate interest in the new directions made possible by SAMEX data. An important interaction will inevitably develop around the SAMEX vector magnetic field data and the ground-based efforts in analyses of Stokes polarimetry presently in formation at the National Solar Observatory, the High Altitude Observatory, the Johns Hopkins University, and the Institute for Astronomy at the University of Hawaii. At MSFC, the SAMEX data will enrich an ongoing program in vector magnetic field observations and analysis. Studies of coronal structure and activity and solar wind generation will benefit immensely from the ability to observe the corona against the disk. Internationally, SAMEX will open new possibilities for collaborative programs in high-resolution studies with the planned Large Earth-Based Solar Telescope (LEST).

Finally, a subtle but highly important overall impact of SAMEX is that the prediction approach to research in solar activity will serve as a means of forcing new developments in the theory of magnetized plasmas and instabilities. The benefits of these will be felt not only in solar physics but in the broader arena of astrophysics as a whole.

A pre-phase A version of SAMEX has been discussed within DOD as a means of developing the requisite science for solar forecasting techniques. The mission has been approved by Air Force Space Division for the Space Test Program, but without provision for funding the instrumentation at the laboratory level. AFGL and NASA seek to define a mission concept which would accommodate, at reasonable cost, the interests of NOAA, NASA, DOD, and the scientific community, and which would be suitable for future planning as a cost-shared mission. Interagency collaboration would be highly cost effective and would satisfy recommendations by the Committee on Solar-Terrestrial Research regarding coordination between federal agencies in major efforts in solar-terrestrial studies (Solar-Terrestrial Research for the 1980's, National Academy Press, 1981; National Solar-Terrestrial Research Program, National Academy Press, 1984).

In summary, strong arguments can be put forward for the need to develop and fly the SAMEX mission. In anticipating the eventual recognition of this need by those federal agencies that would benefit from such a mission, AFGL has promoted the development of initial design concepts for the core instruments of SAMEX. This developmental work has been done by teams who are recognized for their expertise in the various scientific and technical areas encompassed by the SAMEX instrumentation. By far the most complex of the SAMEX instruments is the vector magnetograph. This complexity has necessitated a thorough pre-phase A study to develop a complete design for an instrument to measure all three components of the magnetic field on the surface of the Sun, and to make these measurements over a 3-year time span from an unmanned satellite. The results of this in-depth study comprise the material presented in this report which is a joint effort of the USAF Air Force Geophysics Laboratory and the NASA Marshall Space Flight Center.

2. A Space-Based Solar Vector Magnetograph

The origin of the magnetic fields in stars remains one of the major unsolved problems in astrophysics. Current models hypothesize the interaction of convective and rotational motions, producing a dynamo which is located at the lowest level of the convection zone (Parker, 1975). The magnetic flux then ascends to the surface through the action of magnetic buoyancy (Parker, 1979). The subsequent organization and development of surface magnetic fields to form active regions, coronal structures, and ultimately flares, remains a central theme in solar research (Svestka, 1976; Van Hoven, 1980; Orrall, 1981; Priest, 1982) and

has been the focus of extensive work in both theoretical and observational studies for several decades (for a review see Jordan, S., 1981, The Sun as a Star, NASA SP-450).

Because the solar magnetic field does play such a basic role in all aspects of solar activity, the observation of these magnetic fields is a fundamental part of all solar research. Historically, much emphasis has been placed on observations of only one component of the solar magnetic field, the line-of-sight component (commonly referred to as the longitudinal component), and its magnitude and configuration have been the fundamental data upon which the theories of sunspot formation, flares, and predictions of solar activity have been based.

More recently, observations of all three components of the solar surface magnetic field have yielded important new information concerning the configuration of this field in flare-producing active regions. Analysis of these observations shows that the field is almost always in a highly non-potential or stressed state at flare sites, indicating the presence of electric currents at the locations of major solar activity. These signatures of stressed fields have recently been used to predict locations of possible flaring. Yet the degree of success in identifying active regions that will produce flares, as compared with the much poorer success in predicting the time of the flare's occurrence, is ample testimony that the conditions leading to activity are empirically recognizable but not quantitatively understood.

Considerable progress in understanding how preflare magnetic energy is built up and stored through stressing the magnetic field has come from observations of the vector magnetic field, photospheric dynamics, and the geometry of fibril structures of the chromosphere. Research in these areas has developed in parallel with theoretical studies which have shown how free energy can be stored in force-free coronal magnetic fields that are subjected to shearing motions at the photospheric level. The explosive release of this stored free energy in the form of a solar flare is thought to occur through the development of some type of instability in the stressed magnetic field. The present interpretation is that these fields become unstable, break, reconnect, and simplify, thereby releasing energy to power the flare.

The instabilities leading to flares are suspected to occur several minutes to several tens of minutes prior to the impulsive release of the flare energy, and several well-known flare precursors (early indicators) are associated with this pre-impulsive phase. Filament eruptions form an integral part of many, and perhaps all, flares and can be detected in their early stages by outward motions seen in H-alpha observations. The study of the precursor filament eruption, which often leads to coronal mass ejections, has shown the importance of magnetic instabilities in the very earliest stages of the flare process.

Other phenomena, including enhancement of soft x-ray emission and changes in the degree of polarization of microwave emission, occur within the same time frame and are suggestive of an overall picture of preflare magnetic fields undergoing irreversible changes leading to the flare. Clearly, an understanding of the flare onset depends on detailed information about the structure of the magnetic field at all levels of the atmosphere surrounding an active region.

Despite the recent progress in identifying flare precursors and flare activity with sites of stressed magnetic fields, the long-sought, unequivocal prediction of the flare's location, magnitude, and time of onset remains elusive. The failure is part observational and part theoretical: imprecise knowledge of the plasmas and fields in the active region, and insufficiently developed theory of instabilities and reconnection processes. Fundamental questions remain even with the qualitative scenario of energy storage and release in stressed magnetic fields. Why, for example, do obvious conditions of field stress not always lead to flares, and how, in these cases, is the energy dissipated without producing readily detectable thermal signatures and bulk motions?

If progress in answering these and other questions is to be made from the standpoint of observations of the magnetic field, these observations will have to be made with higher spatial resolution, higher magnetic sensitivity, and over more continuous time intervals than are now available from measurements of the vector field from the ground. Based on what we know about the limitations imposed on magnetograph observations from the ground, it is apparent that operation above the Earth's atmosphere is necessary if these objectives of greater sensitivity and resolution are to be met. This conclusion is based on considerations of the limits placed on a magnetograph's sensitivity by the effects that result from looking through the Earth's atmosphere. Natural fluctuations in the Earth's atmosphere occur on time scales that are short compared to the time needed to measure the vector magnetic field over an average sized active region and, as a result, the degradation in image quality produced by these atmospheric changes will be painfully evident in any ground-based observations. Thus, fluctuations in the Earth's atmosphere place limits on the sensitivity and spatial resolution that can be achieved from the ground, and these limits are above the levels needed for meaningful investigations to advance our knowledge of the solar magnetic field and how it interacts with the solar plasma to produce solar activity.

If a magnetograph must be placed in space to avoid the degradation of measurements produced by the Earth's atmosphere, why must it be a complex vector magnetograph instead of a simpler instrument that measures just the line-of-sight component of the magnetic field? Research carried out with the MSFC vector magnetograph and other vector instruments supplies the

unequivocal answer: the line-of-sight field gives absolutely no information on the non-potential or stressed state of the measured field; that information comes entirely from the transverse component of the vector field. This important relation between the transverse component and the identification of stress in the field is brought out more fully in the discussions of the next section on the scientific basis for a space-based vector magnetograph. Thus, the instrument designed to advance our understanding of how the stressed magnetic field controls and drives solar activity and thus improve techniques for predicting this activity must measure all three components of the solar magnetic field.

It will also be necessary to determine the configuration of the magnetic field above the surface where the vector magnetograph measures the field so that the three-dimensional characteristics of the field can be determined. For this reason an H-alpha telescope will be a necessary auxiliary instrument to be flown with the vector magnetograph. Although H-alpha filtergrams can give no quantitative measurements of the magnetic field, the fibrils and filaments seen in the H-alpha images trace the configuration of the magnetic field in the solar chromosphere. The H-alpha instrument will also provide visual evidence of filament motions and eruptions, and the onset of flares. Thus the combined instrument package of a photospheric vector magnetograph and an H-alpha filtergraph will allow us to study the solar magnetic field from the surface of the Sun up to the transition region with unprecedented resolution and accuracy.

3. Scientific Basis and Objectives

The interaction of magnetic fields and plasmas is a process that is common throughout the universe and is especially relevant to the more dynamic, high-energy phenomena that are observed. In astrophysics, examples include galactic and stellar jets, accretion disks, gamma-ray bursts, pulsars, stellar flares, and stellar activity cycles. Processes controlled by the interaction of field and plasma are encountered in space plasma physics in the study of planetary magnetospheres - the generation of geomagnetic substorms and disconnections of comet tails are some examples. On the Sun, we see these processes manifested in the creation of the magnetic cycle, active regions, flares, and coronal mass ejections.

Our Sun provides a laboratory where we can study the interactions of field and plasma at close range. In this solar laboratory, one of the most intriguing manifestations of a dynamic magnetized plasma is a solar flare. This explosive phenomenon results from processes that are thought to be operative in many high-energy astrophysical events, so that understanding the flare process has widespread implications in astrophysics and space science. To understand the flare phenomenon we must understand the processes by which the energy of a flare is built up, stored, and then released explosively.

The solar magnetic field is the key to gaining such an understanding, and the SAMEX vector magnetograph has the specific objective of measuring the magnetic vector field with unprecedented accuracy and spatial resolution over long periods of time. By surmounting the previous deficiencies in observations of solar magnetic fields - poor spatial resolution, insufficiently sensitive measurements of the magnetic field, our inability to observe continuously over periods longer than several hours, lack of simultaneous observations at different heights, atmospheric distortions - this instrument, along with the other SAMEX instruments, will reveal the processes through which the solar magnetic field generates and releases the energy observed in flares. While the study of the buildup, storage, and release of the magnetic energy in flares is our primary goal, we expect to reap other scientific dividends as well from such long-term monitoring of solar magnetic fields over days, weeks, months, and years. And, as with any new space instrumentation, the Sun is certain to yield major surprises that will lead to new scientific insight about the physics of the Sun. In addition, by gaining an understanding of the processes of flare buildup, we will enhance present-day capabilities for predicting the occurrence of flares and their potential effects on the solar-terrestrial environment.

3.1 Flare Energy Buildup and Storage

The study of flare energy buildup, storage, and release is the central interest and theme of the SAMEX mission. As such it provides the missing link in the study of flares in the solar laboratory. Previous space experiments have concentrated on studying the phenomena that occur after the flare begins, on the flare products, and on the distribution of the energy among the various observed modes of particle acceleration, mass motions, and heating. Space missions presently planned to study solar flares also are composed of instruments to observe the flare products. Thus the SAMEX experiments will complement these missions by investigating how the flare region becomes charged with free energy before the flare erupts, and how the preflare state is destabilized to flare.

The main questions to be addressed by the SAMEX experiments are the following:

1. What is the preflare configuration of the magnetic field?
2. How is the stressed configuration of the preflare field formed?
3. How is the field destabilized?

To adequately address these three fundamental questions, we must be able to measure the vector magnetic field in the photosphere, and to describe the three-dimensional characteristics of this field in the chromosphere and corona by using

tracers of the field as seen in H-alpha and x-ray images. This will require long-duration, continuous observation with instruments of unprecedented sensitivity and high spatial resolution.

3.1.1 Configuration of the preflare magnetic field

The premise that magnetic fields play a dominant role in solar flares has been a well-entrenched theme over the past decade. Solar scientists acknowledge that the free energy necessary to fuel a solar flare must come from the magnetic field (Van Hoven et al., 1980), and to provide this free energy, the fields in flare-productive active regions must therefore be non-potential. A common characteristic of non-potential active regions is a strongly "sheared" field configuration, where we use the term shear in an observational sense to denote the degree to which the transverse (to the line-of-sight) component of the magnetic field at the magnetic inversion line is aligned more parallel than perpendicular across the inversion line. Specific examples of sheared fields associated with flares come from indirect observations of fibril and filament structures seen in H-alpha, EUV, and soft x-ray images which trace the chromospheric and coronal fields (Beckers, 1971; Krieger et al., 1971).

Observations with instruments that measure all three components of the photospheric magnetic field show directly many striking examples of magnetic shear near the polarity inversion line (Moreton and Severny, 1968; Harvey et al., 1971; Krall et al., 1982; Neidig et al., 1986). Such an example is seen in Figure 2 which shows the vector magnetic field of a small bipolar region within a larger area of activity (Boulder region 2372 in April 1980) that was the site of a series of major flares. These data were acquired with the MSFC vector magnetograph system (Hagyard et al., 1982) that was developed by the solar group at MSFC; the MSFC instrument is the prototype for the SAMEX vector magnetograph.

From these direct measurements of magnetic shear obtained with vector magnetographs, convincing evidence for the correlation of photospheric magnetic shear with the occurrence of flares has been found. In a statistical study, Smith and Hagyard (unpublished) have found a distinct preference for high flare productivity and major flares in areas where significant magnetic shear is observed. In evolving regions they found an increase in shear is accompanied by an increase in flare frequency and magnitude, and decreasing shear is commonly accompanied by a decrease in flare production.

Recently quantitative work has been carried out relating the degree of magnetic shear with flare sites. Hagyard et al. (1984b) quantified the concept of shear by defining a parameter, the angular shear. This quantity is defined as the difference, at the photosphere, between the azimuths of a potential field and the observed field, where the potential field satisfies the boundary conditions provided by the observed line-of-sight

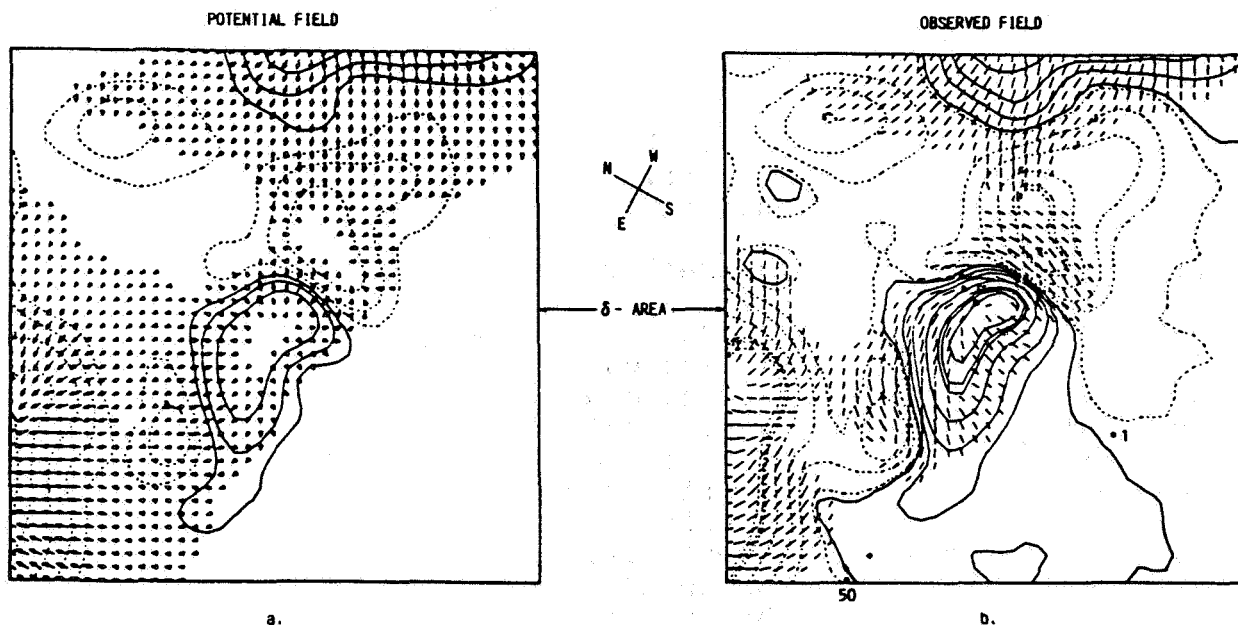


Figure 2. An example of sheared photospheric magnetic fields in an active region from observations with the MSFC vector magnetograph. Solid (dashed) curves represent positive (negative) contours of the line-of-sight magnetic field. The directed line segments represent the strength (by length) and orientation (by direction) of the transverse component of the magnetic field. The field of view is $1.67' \times 1.67'$. (a) The potential magnetic field calculated from the observed line-of-sight field. (b) Observed magnetic field. Note the areas along the magnetic inversion line of the central bipole where the azimuth of the field deviates significantly from the potential field's azimuth. These areas were sites of flare kernels in the flares that erupted in this region.

ORIGINAL PAGE IS
OF POOR QUALITY

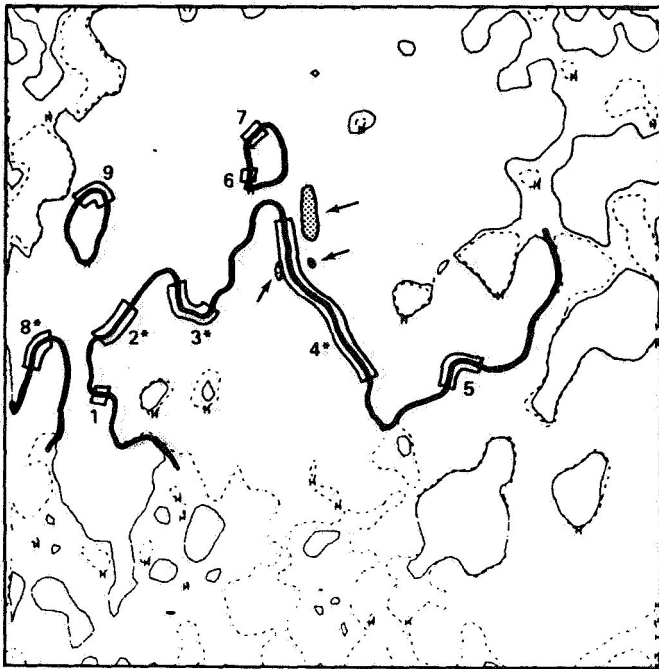
field. Analyzing the magnetic field at the polarity inversion line of the active region shown in Figure 2, they found that, although the field was sheared along most of the length of this inversion line, flares erupted just where the shear and the field intensity attained their maximum values. In a subsequent study of two more flare-producing regions, Hagyard and Rabin (1986) found similar results, namely that the flare onsets always coincided with a local maximum of the angular shear and with areas of strong field intensity. One of the regions studied is seen in Figure 3 which shows the areas of strong shear along the magnetic inversion line and the flare kernels of the white-light flare that erupted in this region (Boulder region 4474, April 24, 1984). While there were extensive areas of strong shear throughout this region (marked by the rectangles in Figure 3a), the shear attained its maximum value at the point of the inversion line bracketed by the flare kernels (shown by the arrows).

As a result of these quantitative studies, Hagyard and Rabin (1986) were able to draw some general conclusions concerning the magnetic field at the sites of flare onset. First, the location of flare onset coincides with a local maximum of the angular shear; this suggests that flares initiate at the exact location where the local field is stressed the most. Second, flares erupt where the angular shear is at least 80° and the field intensity is more than 1000 G. Third, the high angular shear and strong fields prevail over an extended length of the inversion line, on the order of 5000-8000 km or more.

These properties of the highly stressed fields at flare sites are really the photospheric signature of the electric currents flowing in the surrounding solar atmosphere, and quantitative vector magnetograms are the only method available to obtain a measure of these currents. The vector magnetic field of the April 1980 region has been analyzed to determine the nature of the currents flowing in the areas where the flares initiated. In this study, the normal component of electric current density J_z crossing the photosphere was derived from the observed transverse magnetic field. The results showed that, outside of the large sunspots, the maximum concentrations of these vertical currents occurred at sites of flare initiation. This correlation can be seen in Figure 4, especially at points A, B, and E. The lack of measurable current densities at points C and D may be due to the high noise level of the ground-based data, in this case, on the order of $50 \times 10^{-4} \text{ A m}^{-2}$. Weaker currents at points C and D thus would not show up since the lowest contour level was at the noise limit.

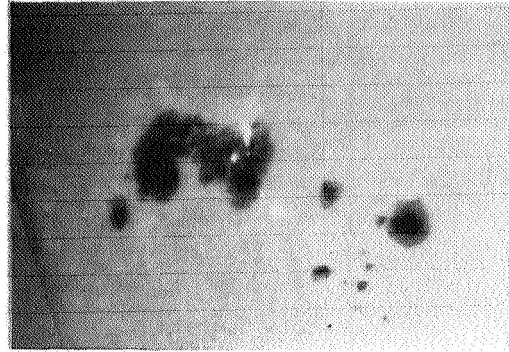
The distribution of current density shown in Figure 4 suggests that currents are flowing out of the area of positive magnetic polarity and across the inversion line into the two areas of negative polarity (Figure 2). This interpretation is supported by a calculation of the "source field" using the technique developed by Hagyard et al. (1981). This source field

AREAS OF STRONG SHEAR ALONG NEUTRAL LINE



a

INITIAL FLARE POINTS



b

Figure 3. (a) Areas of strong shear along an active region's magnetic inversion line. The solid (dashed) curves represent positive (negative) contours at the 5G level. The darker curves outline the main magnetic inversion lines of the active region. Rectangular areas denote regions where the angular shear is greater than 70° . Those marked with asterisks denote areas where the angular shear exceeds 85° and the field intensity is greater than 1000 G. The field of view of this MSFC magnetogram is $6' \times 6'$. (b) Initial flare kernels seen in white-light observations obtained at the Tokyo Astronomical Observatory; these are sketched in (a) as stippled areas marked by the arrows. The angular shear along the inversion line in area 4 reached a maximum value of 90° at the point between the two smaller flare kernels.

ORIGINAL PAGE IS
OF POOR QUALITY

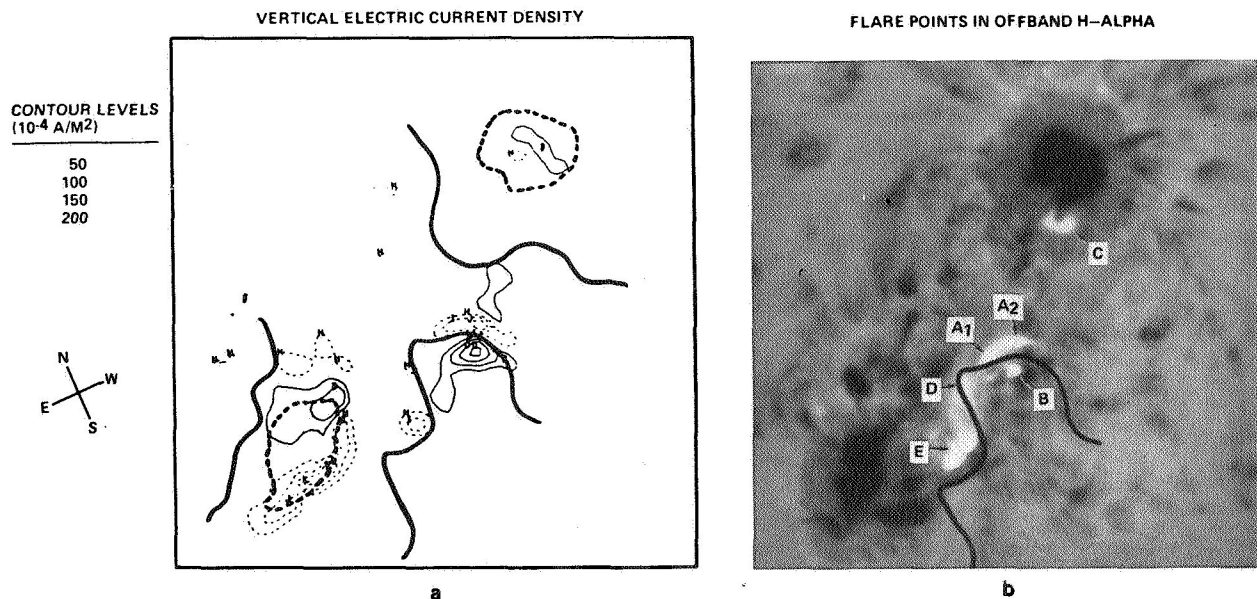


Figure 4. Relation between currents and kernels. (a) Distribution of the vertical electric current density over an active region. The currents are derived from measurements of the vector magnetic field obtained with the MSFC vector magnetograph. Solid (dashed) curves represent contours of current densities flowing out of (into) the photosphere. The darker curves outline the major inversion lines; the darker dashed curves outline the umbral areas of the two major sunspots in the 5' x 5' field of view of the active region. (b) Flare points seen in off-band H-alpha images obtained at the Ottawa River Solar Observatory. The central inversion line of (a) has been superimposed to aid in correlating the currents with the flare kernels. Outside of the two main umbral areas, one can see that the major concentrations of currents coincide with the flare points.

ORIGINAL PAGE IS
OF POOR QUALITY.

is the magnetic field produced in the photosphere by the electric currents above the photosphere and it can be calculated from the observed photospheric vector magnetic field. For the April 1980 active region the calculation shows the source field is comparable to the field produced by two arcades of currents crossing the inversion line with orientations that would align them with the chromospheric field inferred from H-alpha fibrils.

The close association of large angular shear, strong fields, and electric currents with the sites of initial energy release in flares implies that these are all basic ingredients in the buildup and release of flare energy. However, we are a long way from completely understanding their respective roles. For instance, the study by Hagyard and Rabin (1986) provided examples of areas along the inversion line where the angular shear was greater than 80° , the field intensity exceeded 1000 G, and strong electric currents were present, but where there were no flares (regions 2,3 and 8 in Figure 3a are good examples). Consequently a prime objective of the SAMEX vector magnetograph will be to study the configuration of the preflare magnetic field and develop an understanding of the difference between non-potential fields that flare and those that don't. As we shall demonstrate in the following paragraphs, to carry out this fundamentally important scientific objective will require four essential instrumental characteristics of the SAMEX magnetograph: the highest possible magnetic sensitivity, high spatial resolution, a sufficiently large field of view, and the best possible temporal resolution, all needed to resolve and detect the fine-scale structure and dynamics of the magnetic field where these differences may be hidden. (It will also require the information obtained from the other two instruments that shows the configuration of the field from the photosphere into the corona so we can see where in the three-dimensional configuration of the field these differences arise.) The arguments for these four instrumental specifications are straightforward and compelling:

(1) The magnetic sensitivity of the instrument must be as high as modern technology allows to measure the transverse component of the magnetic field with the greatest accuracy possible, since accurate calculations of electric currents and angular shear depend on how precisely we measure the transverse component of the magnetic field. The lowest resolvable electric current flowing vertically through the photosphere, I_z^0 , is approximately given by the expression

$$I_z^0 = (4.1 \times 10^7 \text{ A G}^{-1} \text{ arc sec}^{-1}) B_T^0 p_r,$$

where the term B_T^0 represents the noise level in the transverse component B_T , and p_r is the spatial increment (in arc sec) used in the finite-difference approximations for calculating the curl of B (p_r would normally be the pixel resolution of the instrument). The uncertainty in calculations of the vertical currents I_z is given by

$$\delta I_Z = I_Z^O (B_T^O/B_T) .$$

(We distinguish here between the concept of noise level, meaning the lowest measurable value of a physical quantity, and the uncertainty that is associated with any measurement above the noise level.) With a proposed polarimetric sensitivity of 10^{-4} for the SAMEX magnetograph, calculations for B_T^O give values ranging from 35 to 50 G. Less sensitive instruments currently planned or in operation measure polarization to only about 10^{-3} ; B_T^O for these instruments is much higher. From analysis of data obtained with the MSFC vector magnetograph, values for B_T^O have been found to range from 125 to 180 G. Taking the higher figures for B_T^O in both cases (50 and 180 G), the noise level I_Z^O for vertical electric currents measured with the SAMEX instrument will be 5.1×10^8 A, whereas that for the less sensitive polarimeter will be 185×10^8 A, or 36 times as large. In addition, for a transverse field of 100 G, the uncertainty in the calculation of the electric currents will only be 2.6×10^8 A for the SAMEX magnetograph but will be 334×10^8 A for the case of less polarimetric sensitivity. The SAMEX magnetograph will thus measure these currents with an uncertainty reduced by over 2 orders of magnitude.

Uncertainties in calculating angular shear arise from the uncertainty $\delta\phi$ in measuring the azimuth ϕ of the transverse field, and this can be calculated from the relation

$$\delta\phi \approx \frac{1}{2} (B_T^O/B_T)^2 .$$

To understand the role of angular shear in the flare buildup process and to distinguish areas of angular shear that lead to flares from those that don't, we need to observe the buildup of the shear into the critical range (80° to 90°) with an accuracy of at least 10%, i.e., 1° . And we need to be able to see such changes in areas where the field strength is less than the critical value of 1000 G. For a 300 G transverse field, the SAMEX polarimeter could measure the azimuth and thus the angular shear with an uncertainty of less than 1° , whereas the less sensitive instrument could achieve this level of uncertainty only for field strengths greater than 1000 G. Clearly, we need a polarimetric sensitivity of 10^{-4} .

We also need to measure the magnitudes of the magnetic field as accurately as possible to place stringent conditions on the relative strengths of the magnetic field in flaring versus non-flaring areas. For the same 100 G transverse field, the uncertainty in the measurement with the SAMEX polarimeter is calculated to be 12 G; that for the less sensitive instrument is 160 G, an order of magnitude worse.

(2) The spatial resolution must be at least 0.5". This degree of spatial resolution will reduce the errors in measuring the line-of-sight component B_L , B_T , and ϕ , and in calculating J_Z by more nearly resolving the individual flux elements. It will

also provide a look at the fine-scale structures of the angular shear and currents and how these might be related to the onset of flare activity. Better spatial resolution would, of course, be desirable to clearly resolve the fields outside of sunspots that consist of subarcsecond kilogauss elements (Stenflo, 1973). However, the spatial resolution of 0.5" will be sufficient to measure the magnetic field of most features observed in active regions where flares occur, and to measure the 0.1"-0.4" network fields, we can use the technique of flux tube diagnostics (Stenflo, 1973). This technique imposes some requirements on the spectral range of the SAMEX magnetograph, but the spectral range should be limited as far as possible for technical reasons (e.g., thermal control, design of waveplates, and simplicity).

(3) The field of view must be large enough to completely encompass the magnetic field of most active regions. This condition is needed to ensure that there are no strong sources of the active region's magnetic field outside the boundaries of the magnetogram so that the boundary conditions for potential field calculations are satisfied as nearly as possible. Errors in the potential field calculation arising from improper boundary conditions will cause errors in calculating the angular shear. Also, since we are interested in studying different areas of shear within an active region to determine why some lead to flares and some don't, it is essential that the field of view cover both average-sized active regions and most large ones. To meet the requirement for model calculations, a field of view larger than the area of average active regions is necessary, that is, larger than about 3' x 5' (from the Zurich classification shown in Bray and Loughhead, 1964). Very large active regions have dimensions of about 4.5' x 9.0' (classification F). Thus the field of view of 4.3' x 8.5' we propose for the SAMEX magnetograph provides an optimum choice to satisfy these requirements.

(4) The temporal resolution must be sufficiently high that consecutive magnetograms can document the dynamics of the preflare configuration of the magnetic field. Preflare impulsive phenomena are known to occur on time scales of about 10 minutes before the onset of flares so we should expect to see changes in stressed fields on this time scale. Therefore, we need a magnetograph time resolution on the order of a few minutes. In addition, a temporal resolution of about 5 minutes is required to avoid image smearing due to solar rotation and to insure that phenomena traveling horizontally at speeds below the acoustic velocity of 7 km/sec will be resolved.

3.1.2 The formation of non-potential fields

The high-beta plasmas in the photosphere are the drivers of the field evolution in the low-beta coronal plasmas, and thus we look for the sources of shear in the high-beta plasma regime. By far the best-observed mechanism for the formation of shear is sunspot motion. Opposite-polarity spots have been

observed to move in opposite directions tangentially to the magnetic inversion line prior to periods of flare activity (e.g., Gesztelyi et al., 1986). The motions presumably distort or shear the interconnecting magnetic fields and thus build up the magnetic energy that is released in flares.

Sometimes shear can develop from head-on collisions of spots of opposite magnetic polarity rather than from shearing motions, as demonstrated by Gaizauskas and Harvey (1986) in a study of two colliding sunspots in May 1978. Connections of the magnetic fields of these two spots were indicated by the penumbral and H-alpha fibrils linking the pair; these linkages had the appearance of potential configurations until after the fifth day of observations. Then something different happened: the penumbral structures bent along the line of collision indicating the field was sheared. H-alpha structures also changed orientation: their inclination to the inversion line decreased in one area by 30° in 4 hours even though the sunspots did not change direction. Subsequent to these changes in the field, major flares erupted in the area of collision where the field was highly sheared. These data show that the development of the shear from the collision was an essential prerequisite for major flare activity. But the lack of high resolution vector magnetograms for this event prevents us from knowing the precise process leading to the "bending" of the penumbral filaments and shearing of the previously unstressed field.

These observations provide an indication of the accuracy and temporal resolution needed to measure changes in the angular shear, in this case, the buildup of the shear. The change of 30° in 4 hours gives an average rate of change of 1° every 10 minutes. To detect such changes, the uncertainty in the measurement of the field's azimuth must be less than 1° . For a polarimetric sensitivity of 10^{-4} , transverse field strengths above 250 G are required to reduce the uncertainty below 1° ; with a sensitivity an order of magnitude worse, fields greater than 950 G are needed. In this latter case the areas where such changes could be detected would be limited to areas of strong transverse fields, i.e., near the umbral region, whereas, with the higher polarimetric sensitivity we propose, these changes could be detected in the weaker fields near the penumbral border at the photosphere where the bending of the field presumably starts. With regard to temporal resolution, since changes of 1° may occur in a time interval of about 10 minutes, we must be able to acquire magnetograms at a higher cadence, at least on the order of one every 5 minutes.

Perhaps it is only the easy association that can be made between spot motions and shear formation that makes this mechanism attractive. For clearly, there are observed areas of shear that were not formed by spot motions (e.g., Athay et al., 1985). In these cases other mechanisms may be operative, such as large-scale, sheared plasma flows generated by the convergence of giant circulation cells (Athay et al., 1986), flux emergence

(Zirin, 1983), and flux submergence (Rabin et al., 1984). Recently, S. F. Martin (1986) has studied the development of shear by observing the formation of filaments. The presence of filaments has been used for some time as a reliable indicator of the existence of sheared magnetic fields in the chromosphere and above. In a study of the formation of two small filaments, observations by Martin showed that the filaments formed at sites either adjacent to or coincident with sites where cancellation of flux had taken place either continuously or intermittently for an interval of hours to days before the filaments formed. The cancellation presumably results from magnetic reconnection which changes the topology of the field, leading to a configuration suitable for the formation of filaments, i.e., a sheared configuration.

While all these processes are viable mechanisms for forming stressed, non-potential fields in active regions, the evidence supporting their respective roles in shearing the fields is mostly circumstantial, based on observations of causative processes and not on direct observations of the magnetic field becoming increasingly stressed. With the SAMEX vector magnetograph, such direct observations will be possible. For example, we can only infer that shear develops as spots move tangent to the magnetic inversion line. This association can be verified by observing the magnetic and velocity fields of evolving active regions at the photospheric level where the photospheric plasma carries the field as it moves. However, in such a study, it will be necessary to have a field of view large enough to image the entire active region so that areas where shear starts to build up or increase can be identified. The proposed field of view of the SAMEX instrument will provide the necessary areal coverage.

To see subtle changes in the field as it becomes increasingly sheared will require the high magnetic sensitivity and spatial resolution of the proposed vector magnetograph. The initial process of the buildup of shear may involve small changes in the azimuth in areas of weak transverse fields. For example, if an initially unsheared loop whose footpoints are separated by 9000 km has its footpoints sheared at 0.1 km/sec, in 10 minutes the angular shear will only be about 1° . This will be observable (as shown previously) in areas of weak transverse fields (about 250 G) if the polarimetric sensitivity is 10^{-4} . Observing these subtle changes will also require high spatial resolution, which also increases magnetic sensitivity.

With the high spatial resolutions of the H-alpha and x-ray telescopes we will be able to observe the response of the magnetic fields higher in the solar atmosphere to these movements of the photospheric footpoints. Perhaps these responses will prove to be different for areas of highly stressed fields that flare and those that don't. Our experiment will directly observe the driving photospheric flows, since the vector magnetograph will measure photospheric velocities from Dopplergrams and observations of proper motions.

With its unique combination of high sensitivity and spatial resolution and with the capability for three-dimensional spatial observations and uninterrupted monitoring, this experiment will also allow us to directly observe flux emergence, cancellation, and submergence, and thus test current hypotheses describing their respective roles in stressing active-region fields.

In studying the flare buildup process through observations of the magnetic field, we will also concentrate on how the field intensity builds up at flare sites, that is, does it occur by emergence or by coalescence? And here again we will look for differences between regions that are flare-productive and those that aren't.

3.1.3 The free energy stored in stressed fields

The storage of free magnetic energy in stressed magnetic fields has been studied using self-consistent MHD models of magnetic loops undergoing shearing motion. For example, Wu et al. (1984) have demonstrated that more than enough magnetic energy is built up by such a process to provide the energy released in a typical solar flare. The energy accumulates as the shear is increased, with the growth rate roughly proportional to the shearing speed, to the length of the affected inversion line, and to the initial strength of the magnetic field. The highest concentration of energy is located near the inversion line and this concentration is more pronounced if the initial magnetic field is already in a non-potential configuration. These results are completely consistent with the observational data relating locations of energy release with the inversion line and areas of maximum shear.

Use of ground-based observations of magnetic fields in active regions to confirm or confront the quantitative results of such models is open to question for two reasons. First, an exact calculation requires knowledge of the total vector field throughout the volume where the energy is to be calculated. Since this information is not usually available, some assumptions have to be made about the field. For instance, we might assume that the field is force-free so that the currents must be parallel to the field; this conditions can be stated mathematically as $\nabla \times \mathbf{B} = \alpha \mathbf{B}$. Presently only the case of linear force-free fields can be treated but this means that the parameter alpha must be constant over the region of the force-free calculation. But alpha actually prescribes the sense and amount of shear in the field, and we know from experience that these are not the same throughout an active region (Levine, 1976; Krall et al., 1982; Gary et al., 1987). Consequently, the calculation of energy by this method is very restricted in its application. Under less restrictive assumptions, Low (1982) has shown that the free magnetic energy for any force-free field (linear or non-linear) of limited spatial extent can be calculated from the vector field measured over the photospheric surface.

Gary et al. (1987) have used this formalism of Low with measurements of the vector field obtained with the MSFC vector magnetograph to calculate the free energy stored in an active region observed by the Solar Maximum Mission experiments in September 1980. Their calculations highlight the second reason it is difficult to use ground-based measurements to calculate the energy, namely, the noise in the measurements. Gary et al. calculated the total non-potential magnetic energy stored in the region and showed it was of order 10^{32} ergs, which was about 3 sigma above the noise level. They thus concluded that the uncertainty in vector field measurements will have to be substantially reduced in order to confidently measure the buildup and release of magnetic energy in any but the very largest flares.

To determine the levels to which uncertainties in measurements with the SAMEX instrument must be reduced to observe significant changes in magnetic energy, we can estimate the uncertainty δE in the calculation of total magnetic energy E and compare it to the energy E_F released as free magnetic energy in different classes of flares. To observe changes in the calculated magnetic energy that are on the order of the free magnetic energy, the uncertainty δE must be much smaller than E_F . In Table 1 we show estimates of E_F for different classifications of flares.

Table 1. Free magnetic energy released in flares of different x-ray classifications.

Class	Free energy E_F (ergs)
X	10^{32}
M	10^{31}
C	10^{30}

To estimate the uncertainty δE in calculating the total magnetic energy from measurements of the vector field, we use the expression for the magnetic energy of a force-free field discussed by Low (1982b). In that formulation the total magnetic energy E can be calculated from the following surface integral:

$$E = \frac{1}{\mu_0} \int_0^{L_x} \int_0^{L_y} B_z (x B_x + y B_y) dx dy .$$

This integral can be approximated by an expression of the form

$$E \approx \frac{L^3}{2\mu_0} [\langle B_x B_z \rangle + \langle B_y B_z \rangle] ,$$

where L^3 is the volume in which the energy E is calculated, and $\langle B_x B_z \rangle$ and $\langle B_y B_z \rangle$ are products of the field components averaged over the xy surface at $z = 0$. Using this approximation, an expression can be developed for the uncertainty δE in calculating the energy due to uncertainties in the measured values of the field components; the following approximate form was derived:

$$\delta E \approx \frac{\sqrt{2}L^3}{4\mu_0} \frac{(B_T^0)^2}{N^{1/2}} ,$$

where N is the number of discrete points over the surface $z = 0$ at which the field is measured. In Table 2, we compare the uncertainty δE in calculating the total magnetic energy from measurements with the MSFC vector magnetograph and the SAMEX instrument. In both cases we used both the maximum and minimum expected values for B_T^0 .

Table 2. Uncertainties in calculating total magnetic energy with two instruments of different sensitivities.

System	L	N	B_T^0 (G)	δE (ergs)
MSFC	6' x 6' x 6'	128 x 128	125	0.6×10^{32}
			180	1.3×10^{32}
SAMEX	8.5' x 4.3' x 8.5'	1024 x 2048	35	0.6×10^{30}
			50	1.3×10^{30}

These estimations illustrate very clearly the reason why changes in magnetic energy might not be observed in even the largest flares ($E_F \sim 10^{32}$ ergs) using magnetographs with low sensitivity: the uncertainties are on the order of the energy released. On the other hand, these figures indicate that, with the sensitivity we are aiming for with the SAMEX vector magnetograph, we can use the Low method to measure the variation of free magnetic energy at the level of most X- and M-class flares as a function of time.

Accurate calculation of changes in free magnetic energy also depends very much on the field of view of the observations since this affects the boundary conditions of the numerical extrapolation used to calculate the energy. The basic force-free field equations require that there is zero net flux over the domain of the calculation. For magnetograms which have a non-zero net flux because the field of view does not encompass all the flux related to the active region, evaluation of the magnetic energy using a force-free extrapolative method requires that all values of the field used in the calculation be adjusted by subtracting the average net flux from each value. If this average flux is greater than the uncertainty in the measured field, it is doubtful whether any calculation of energy would be very meaningful.

The effects of boundary conditions on different extrapolative techniques have been investigated by Seehafer (1982). Besides demonstrating differences in the structure of field lines, the height to which field lines extend, the number of field lines leaving the domain of calculation, and the strength of the field as a function of height, Seehafer showed that the boundary conditions (and thus indirectly the field of view of the observations) have a major influence on calculations of the magnetic free energy with differences amounting to an order of magnitude occurring when boundary conditions are changed.

Thus, in computing the free magnetic energy, it is extremely important to measure the vector magnetic field not only with the best possible sensitivity, but also over a sufficiently large field of view to avoid the inaccuracies induced by improper boundary conditions. The field of view proposed for the SAMEX instrument is large enough for this purpose.

Since the Low method assumes that the photospheric field is force-free, the resulting calculation for the energy may be inexact because the force-free condition is likely to be a poor approximation in the photosphere. However, Low (1985) has developed a quantitative test for whether the force-free state exists in the volume above an active region using measurements of the vector magnetic field of the active region in the photosphere. Thus, observations with the SAMEX vector magnetograph will simultaneously provide the data needed to calculate the maximum energy available for release in flares in force-free fields and allow a quantitative estimation of how close the fields are to being force-free.

3.1.4 Modeling of non-potential fields

The data from the SAMEX vector magnetograph will be used in computational models that have been developed to help us describe the overall configuration of the solar field. Scientists at MSFC have operational computer programs for the following potential and force-free methods: the Schmidt-Semel-Harvey potential method (Harvey, 1966), the Teuber potential

method (Teuber et al., 1977), the Nakagawa-Raadu-Welck force-free formulation (Nakagawa and Raadu, 1972), and the Alissandrakis force-free formulation (Alissandrakis, 1981). We are also investigating two methods for calculating non-linear force-free fields using vector magnetograph data as input. Wu et al. (1985a) have developed a numerical method using vector data as the initial boundary condition in an implicit forward-difference method with a backward-looking verification procedure. The second method, developed by Pridmore-Brown (1981), is a variational method. It uses a Galerkin technique in which the difference between the observed field and a potential field satisfying the boundary condition of the observed line-of-sight field is represented by a linear superposition of finite basis functions chosen from a complete trigonometric set. The coefficients of the basis functions are determined by minimizing both the Lorentz force in the volume above the photospheric boundary and the deviations between the azimuths of the observed and calculated field. A computer algorithm has been developed for this method and verified for several test cases using analytical models.

Using these computational models, we will be able to construct extrapolated, three-dimensional fields for comparisons with the fields traced out by the H-alpha and x-ray structures. These comparisons will indicate the degree to which the real solar fields are non-potential and whether the non-potential fields can be classified as force-free or not. This technique will complement the quantitative method of Low (1985) for testing the validity of the force-free approximation. The computational models will also be used to see how the magnetic energy is distributed spatially. We will be able to investigate the relation between the distribution of angular shear and field intensity and the concentration of the free energy, and relate these quantities to flare characteristics.

In proposing to use these computational models we must keep in mind that the validity of the numerical methods depends on several factors. All methods require that certain boundary conditions must be met; this requirement imposes a general condition on the field of view of the SAMEx instrument, namely, that it cover the entire active region's photospheric magnetic field so that the net flux over the region is zero (within the accuracy of the measurement of the field). Accurate calculations require that the field be spatially resolved to the extent possible with a moderately-sized telescope. With the 0.5" spatial resolution proposed for SAMEx and using flux tube diagnostics with appropriate spectral lines, we will be able to adequately resolve even the smallest field elements and thus account accurately for their contributions to the numerical computations. Also, the overall accuracy of these numerical techniques depends on how accurately we measure the magnetic field and thus on the ultimate sensitivity of the polarimeter. As shown previously the magnetic sensitivity drastically affects calculations of free magnetic energy, angular shear, and electric

currents. While we cannot give quantitative examples of how uncertainties in the measured field would affect these numerical models, we would expect that the effects would be similar to those just mentioned and thus provide further justification for the increased magnetic sensitivity that we propose for SAMEX.

3.2 Instability of the Field and Energy Release

Observations of the vector field in the photosphere show that, in the cases studied, flares apparently initiate at the point of a local maximum in angular shear and where the field intensity is greater than about 1000 G (Hagyard et al., 1984b; Hagyard and Rabin, 1986). These results suggest that the field evolves toward a critically stressed configuration, and exceeding this critical state either triggers, or is a necessary condition for triggering, the release of energy in the form of a flare at the site of excess stress. Subsequently the field relaxes to a lower energy state as the energy dissipates in the flare. It is indeed physically plausible that a magnetic configuration can be sheared only so much and that beyond a certain point - a critical value of shear - the field becomes unstable. This idea has theoretical foundations based on the results of studies in which the magnetic shear in a family of static models is successively increased and equilibrium solutions sought (Low, 1977a,b; Birn et al., 1978). It is found that the shear can be increased so far, up to a critical value, beyond which point there are no equilibrium solutions.

Attractive as this scenario seems, it is clear that our picture of what happens during the flare buildup is not complete since we know there are equally stressed fields where flares don't erupt (Hagyard and Rabin, 1986). Either this scenario is basically valid but there are differences between highly stressed regions that flare and those that don't and these differences are too subtle for detection with existing instrumentation, or the presence of strongly stressed photospheric fields is only a necessary but not sufficient condition for triggering a flare and some other condition or perturbation is needed to destabilize the system (De Jager, 1986; Gaizauskas, 1986). The destabilization may be related to reconnection due to emerging flux (Heyvaerts et al., 1977) or evolving flux (Priest, 1985; Machado and Moore, 1986), which may lead to radical changes in the coronal topology in response to the formation of current sheets induced by minor changes at the photospheric level.

With the SAMEX magnetograph and H-alpha and x-ray instruments we can investigate the validity of these two different scenarios for destabilizing the field and releasing the energy. While present theories describing these mechanisms for the primary energy release predict that we will be unable to spatially resolve the area of the energy release, we can learn something about the causative instability by studying the temporal evolution of the resolvable fields. To investigate the flare-triggering scenario of a critically stressed field, we will

observe whether the field continuously evolves towards more and more stressed configurations with a subsequent initiation of a flare as the stress is increased further. We shall see if this evolution occurs slowly or rapidly, and if the transition to instability is a smooth one. Alternatively, to study the second scenario for triggering flares, we will observe whether the field attains a final stressed state and remains in that state until disrupted by the onset of reconnection with newly-emerged flux or by some other destabilizing agent such as localized flows. To unequivocally determine the cause of the instability will require continuous observation and acquisition of magnetograms and H-alpha and x-ray images with high sensitivity and the best possible temporal and spatial resolutions.

To observe the continuous evolution of the field to a critically stressed state, we will need magnetograms of the highest sensitivity to detect the subtle changes that have eluded ground-based observations with polarization sensitivities of only 10^{-3} . While this sensitivity is good enough to see small changes in line-of-sight fields, we are looking for indications of a critically stressed state where flares erupt - at the magnetic inversion line where the field is completely transverse to the line-of-sight. Thus the evidence we are looking for will probably be the small changes in the transverse field that cannot be detected with low sensitivity. Detection of these subtle differences will require that uncertainties in the measurement of the transverse field be significantly below the level of the changes we wish to observe. We can calculate the expected uncertainties δB_T in measuring the magnitude of the transverse component B_T from the following relation:

$$\delta B_T \approx \frac{1}{2} B_T^O (B_T^O / B_T) .$$

As we have indicated, an instrument with the sensitivity proposed for SAMEX will have noise levels B_T^O of 35 to 50 G compared to 125 to 180 G for an instrument with a 10^{-3} sensitivity. In Table 3 the uncertainties in the transverse component for instruments with these two different sensitivities are given as functions of the transverse field strength B_T .

Clearly if we are going to detect changes of a few tens of gauss in the transverse field as the field approaches its critical configuration, uncertainties in the measurements of this component must be even smaller. Only with the SAMEX level of sensitivity will this be possible for reasonable values of field strengths.

Table 3. Uncertainties in the transverse magnetic field for instruments with magnetic sensitivities of 35, 50, 125 and 180 G in the transverse component.

B_T (G)	δB_T (35 G)	δB_T (50 G)	δB_T (125 G)	δB_T (180 G)
20	31 G	62 G	390 G	810 G
50	12	25	156	324
100	6	12	78	162
200	3	6	39	81
300	2	4	26	54
400	1.5	3	20	40
500	1.2	2.5	16	32
750	< 1	1.7	10	22
1000		1.2	8	16
2000		< 1	4	8
3000			2.6	5.4

To determine whether there is a critical value of angular shear just before the onset of flares, we must be able to measure very small changes in the azimuth of the transverse field, on the order of a degree or less. Again, this means that the polarimetric sensitivity must be as high as possible to see small changes in weak transverse fields. To show this we have compared the levels of uncertainty in measuring the azimuth for instruments with sensitivities of 10^{-4} and 10^{-3} ; Table 4 lists these uncertainties for the lowest and highest expected noise levels for both cases: 30 and 50 G, and 125 and 180 G, respectively. To calculate the uncertainties $\delta\phi$ we have used the relation

$$\delta\phi \approx \frac{1}{2} (B_T^O/B_T)^2 .$$

These data indicate that an instrument with a sensitivity of 10^{-4} will be able to detect small changes or differences of a degree in the azimuth of transverse fields of moderate strength. If we want to discover the difference between stressed fields that lead

to flares and those that don't, and if we want to confirm whether or not there is a critical value of angular shear just before flare onset, it will be necessary to observe the azimuth of the field with this accuracy (Hagyard and Rabin, 1986).

Table 4. Uncertainties in measurements of the azimuth of the field.

B_T (G)	$\delta\phi$ (35 G)	$\delta\phi$ (50 G)	$\delta\phi$ (125 G)	$\delta\phi$ (180 G)
20	88°	179°	> 360°	> 360°
50	14	29	179	> 360
100	4	7	45	93
200	1	2	11	23
300	0.4	0.8	5	10
500	0.1	0.3	2	4
750	< 0.1	0.1	0.8	2
1000		< 0.1	0.4	1

On the other hand, to determine if flares are triggered by some destabilizing agent such as emerging or submerging flux, we will need high temporal and spatial resolutions and accurate magnetometry to be able to distinguish flux changes associated with flare triggers from those changes that are associated with the natural evolution of the magnetic field in active regions. The high spatial resolution is necessary to precisely correlate locations of flux changes with flare points observed in the high-resolution H-alpha and x-ray images, thus showing where the initial heating starts and suggesting whether reconnection or dissipation is the source of heating. High time resolution is required for the best possible chances of observing field changes during the ~10-minute time interval just prior to flare onset, when preflare impulsive phenomena occur. And accurate magnetometry is needed so that small changes in relatively weak transverse fields can be detected with some certainty to observe emergence or submergence of the field and to study how new fields interact with existing fields in ways that lead to flare onset and thus differentiate this process from the ongoing evolution of active-region fields.

In looking for flux emergence or submergence, we expect the changes in the field to be seen first in the transverse component (Brants, 1985), and thus we require measurements of the transverse field with the highest possible sensitivity. For example, typical flux changes in the evolution of active regions are on the order of 10^{19} Mx per hour (Heyvaerts et al., 1977; Wallenhorst and Howard, 1982; Rabin, et al., 1984). Flux emerging at this rate over the resolution area of the SAMEX instrument ($0.5'' \times 0.5''$) would result in changes in the transverse field of approximately 600 G in a time interval of 5 minutes which represents an upper limit on the time resolution desired for measuring the transverse component of the vector field. To observe changes on this scale, we would like to measure the transverse field with an uncertainty below 10% of this value, i.e., about 60 G. The data in Table 3 indicate that measurements of this accuracy could be achieved with a SAMEX polarimeter whose noise level is 50 G (the worst case) for all emerging transverse fields greater than about 20 G. However, for the less sensitive instrument (10^{-3} sensitivity), with a noise level of 180 G (worst case), these changes could be detected only for fields greater than about 300 G. Moreover, if the flux emerges at the rate of 10^{19} Mx per hour over the dimensional scale of a pore, i.e., with a radius of approximately 1,000 km, the change in the emerging transverse field would be on the order of 50 G in 10 minutes. To observe changes on this scale, we would need to measure the transverse field with an uncertainty on the order of 5 G. Looking at the data in Table 3, we see that such measurements could be achieved with a SAMEX instrument whose noise level is 35 G (best case) for transverse fields greater than 100 G. With the most optimistic sensitivity for the less sensitive instrument (noise level of 125 G), transverse fields of about 2000 G or more are needed to achieve the desired accuracy. Again it is obvious that we have to go to the higher sensitivity in order to detect flux emergence and submergence in the transverse component on the scale of the field strengths anticipated in these changes of flux.

Flare theories based on the premise that the source of energy released in flares is the magnetic field predict some change in the magnetic field as a result of the flare process. Ground-based observations have so far provided little convincing evidence for such changes because of insufficient cotemporal observations at different heights, because of variable seeing effects that induce observed changes, and because of insufficient sensitivity to detect small changes in the transverse field. What results we do have from ground-based observations show observable changes only in the very large flares (Rust, 1973; Livingston, 1974; Tanaka, 1978). This would imply that changes occur in smaller flares but these changes are too small to be detected with the sensitivity of existing instruments. With the SAMEX instrument, we should be able to see flare-associated changes in the measured vector field in both large- and moderate-sized flares. For example, Rust (1973) observed changes on the order of 400 to 700 G in the line-of-sight component of the

magnetic field in the 3B flare of August 7, 1972. These changes occurred in the penumbral regions of a sunspot where transverse fields are typically 1000 to 1500 G, and they took place in a 90-minute interval. If the rate of change was constant over the 90 minutes, then changes on the order of 45 to 80 G would have occurred over a 10-minute interval. For fields of 1000 G, the data in Table 3 indicate that an instrument with a sensitivity of 10^{-3} (noise levels of 125 to 180 G) might marginally be able to detect these changes, since the uncertainty in the measurements would be on the order of 8 to 16 G. On the other hand, the SAMEX instrument could definitely detect such changes since the uncertainty would be less than 1 G even for the worst case of a 50-G sensitivity.

In another study of flare-related magnetic field changes, Tanaka (1978) observed line-of-sight changes of 10 to 115 G in very weak fields (less than 100 G) in a 2B flare. In these flares, only instruments with sensitivities of 10^{-4} could begin to detect changes of this magnitude in the transverse field. Clearly, if we want to observe changes in the vector magnetic field during flares and thus learn more about the role of the vector field in the buildup and triggering of the flare process, we must go into space with an instrument of the highest polarimetric sensitivity.

In the case of the smallest flares where we may not be able to detect significant changes in the magnetic field, simultaneous observations of photospheric fields and flare-associated changes in coronal x-ray structures will allow us to set upper limits on the amount of free magnetic energy which is dissipated in a given event. This will allow comparison with the observed energy output in the flare with higher accuracy than previously possible. We will also try to determine if the energy release is different in compact and two-ribbon flares. In addition, we will pursue our analysis of highly stressed fields that don't flare to find some differentiating factors that lead to destabilization of the field in the one case but not in others.

3.3 Additional Scientific Dividends

3.3.1 Active regions

Active regions result from emergence of magnetic flux from beneath the photosphere. In the absence of flares, these regions are the brightest and hottest features in the corona. Heating and flaring are most pronounced at those places within an active region where the magnetic field structure is most complex and non-potential. The active region fades in brightness and activity as its magnetic field decays away over a span of weeks. The mechanism through which the field is removed is not settled, but a plausible possibility suggested by recent observational studies and theoretical considerations is that most of the flux disappears by submergence rather than by diffusing into quiet regions (Rabin et al., 1984; Moore and Rabin, 1985).

With the SAMEX complement of instruments, we will explore how coronal magnetic structure and heating in active regions are related to the evolution of the magnetic structure in the chromosphere and photosphere by continuously observing these aspects, tracking selected active regions over the 2-week time span set by the Sun's rotation. Our observations will allow us to investigate the following questions about the magnetic field and its effects in active regions:

- o How do sunspots form? How do they disappear? What is their vector magnetic field? How do their velocity and brightness structures at 0.5" resolution change with time? What is the magnetic field direction and plasma flow in the bright grains and dark fibrils in the penumbra? What is the relation between umbral oscillations and running penumbral waves in the chromosphere and photosphere? Are umbral dots oscillatory?
- o Does the proper motion of clumps of magnetic flux and sunspots match Doppler measurements of cospatial photospheric flow? If not, this would indicate that the field is controlled by subsurface flows that differ from those in the visible layers of the photosphere.
- o How much does the heating of coronal features vary over time spans of minutes, hours, and longer? How does the brightness and variability depend on the vector magnetic field measured in the photosphere?
- o What are the fine-scale features of the photospheric magnetic field, photospheric flow, and coronal structure in emerging flux regions? What is the three-dimensional field structure and how does it affect coronal heating over places where flux disappears in the photosphere? Does the evolution of the coronal and chromospheric magnetic configuration in conjunction with the evolution of the photospheric vector magnetic field and photospheric flow give evidence for or against the removal of magnetic flux by submergence? Is flux cancellation in the photosphere (Martin et al., 1985) accompanied by reconnection above the photosphere? What is the coronal signature of flux cancellation events?

To adequately address these questions, the SAMEX magnetograph must meet the previously specified instrumental requirements on field of view, spatial and temporal resolutions, and sensitivity. The study of active-region magnetic fields demands a field of view large enough to include the entire active region - sunspots and surrounding plage - and the SAMEX field of view of 4.3' x 8.5' meets this condition for all but the largest active regions. The spatial resolution of the instrument must be

high enough to observe the smallest regions of emerging flux, individual granules, intergranule lanes, and filigree within these lanes. Observations with the Solar Optical Universal Polarimeter from Spacelab 2 demonstrated that all these small elements are clearly visible with 0.5" spatial resolution (Title et al., 1986). A temporal resolution of about 5 minutes is required to match the lifetimes of short-lived features such as granules, spicules, and pores, and to observe umbral oscillations and penumbral waves.

The quest for the mechanism(s) through which magnetic flux is removed requires enough magnetic sensitivity to detect fine-scale features involved in flux emergence, submergence, or cancellation. To determine what this level of sensitivity is, consider the observations by Martin et al. (1985) of cancellation of magnetic flux in a decaying active region. They observed a small area of positive field with a maximum field strength of approximately 50 G that disappeared through cancellation with an area of negative field at a typical rate of 2×10^{18} Mx per hour. Over the spatial area resolved by the SAMEX instrument (0.5" x 0.5") this rate results in changes in the magnetic field of 25 G per minute or 125 G over a 5-minute time interval. For a field strength of 50 G, the data in Table 3 show that, under the best conditions, measurements of field changes with the SAMEX instrument would have an uncertainty of about 10 G, which is about 10% of the change (125 G) to be measured and therefore should be resolvable. However, for the less sensitive instrument, the uncertainty would be on the order of 160 G, which is of the order of the changes taking place and therefore these changes could not be resolved.

3.3.2 Quiet-region filaments

In active regions and in quiet regions, magnetic neutral lines across which there is strong magnetic shear are often marked by filaments of chromospheric material threaded and suspended at low coronal heights by the sheared field. In many flares in active regions, the filament serves as a field tracer showing that the sheared field in and around the filament erupts during the onset and impulsive phases of the flare (Moore et al., 1984, 1986). Quiet-region filaments mark neutral lines between opposite-polarity remnants of decayed active regions. These filaments often erupt at the onset of a two-ribbon flare and are ejected from the Sun as the core of a coronal mass ejection. Hence, it is of great interest in connection with flare energy buildup, the flare instability, and coronal mass ejections to determine the magnetic field configuration in and around quiet-region filaments, how this configuration forms and evolves, and how it loses its equilibrium and proceeds to drive the eruption of the sheared field holding the filament (see Figures 5 and 6 and their captions for a description).

The planned SAMEX complement of coronal images, chromospheric images, and photospheric vector magnetograms and

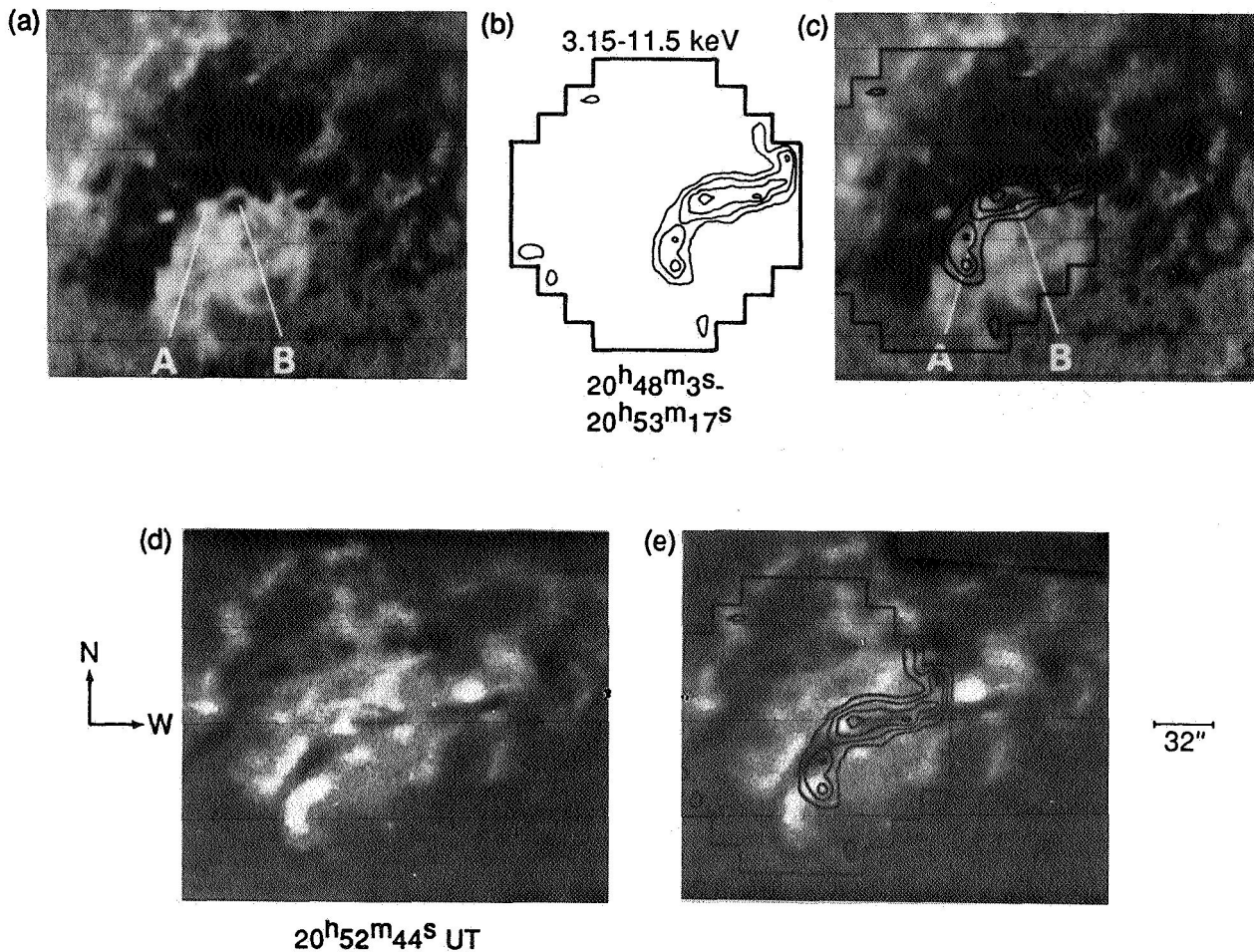


Figure 5. Heating in the sheared field holding the filament at the onset of the filament eruption and large two-ribbon, long-duration flare on May 21, 1980. The heating was detected in soft (3.5 keV) x-rays imaged by the Hard X-Ray Imaging Spectrometer on the Solar Maximum Mission satellite. This onset emission was strongest at the site of emerging magnetic flux (A and B in panels a and c). The magnetograms and H-alpha images are from ground-based observatories and have spatial resolution no better than a few arc seconds. The SAMEX magnetograms and H-alpha images will have an order of magnitude better resolution than these. The x-ray brightness contours shown here were constructed from images with 8" spatial resolution; the SAMEX x-ray images will have 16 times better resolution.

ORIGINAL PAGE IS
OF POOR QUALITY

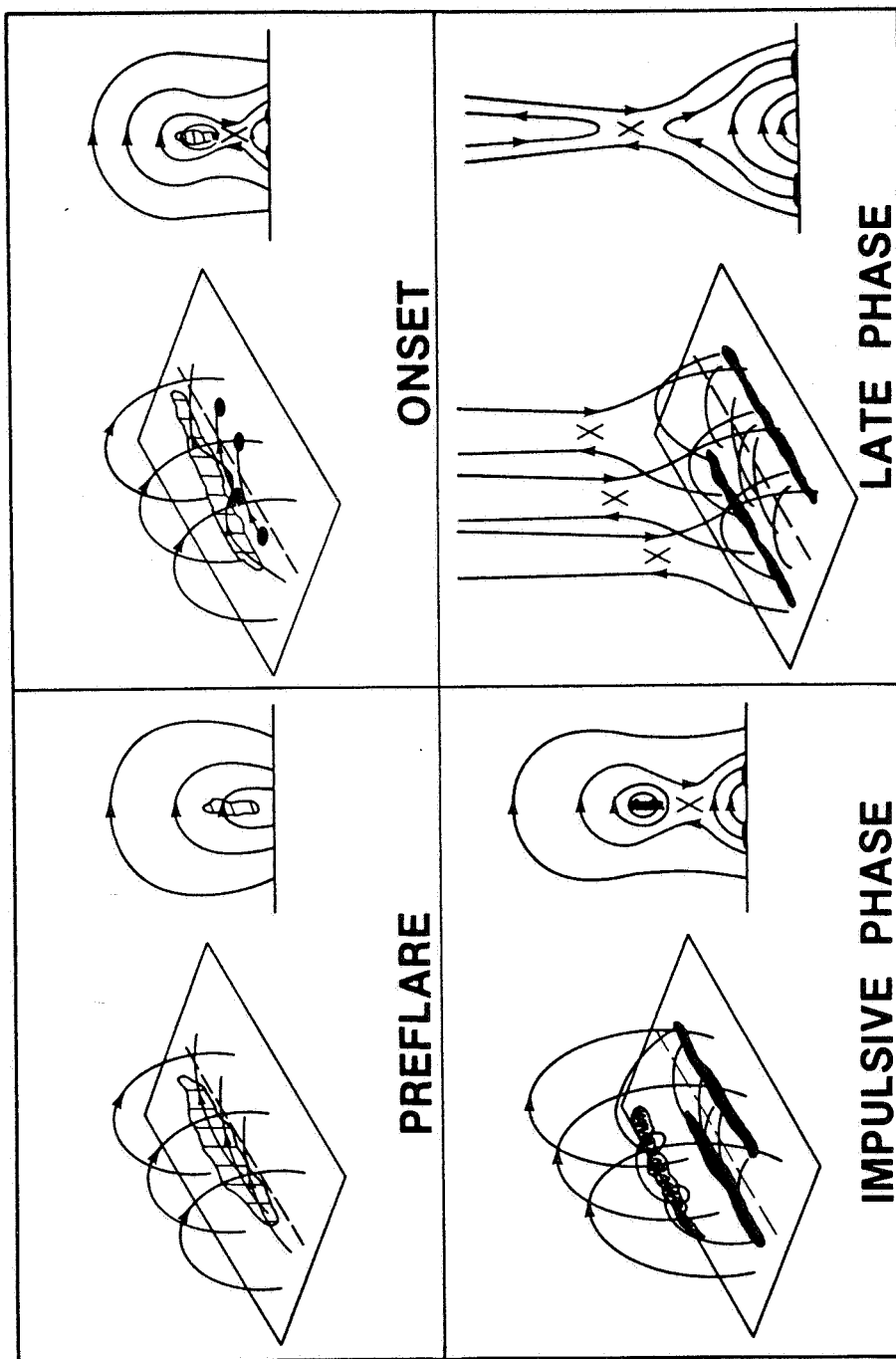


Figure 6. Candidate three-dimensional magnetic field configuration and its change in a filament-eruption flare that produces a coronal mass ejection with the filament at its core. Existing observations of the magnetic structure in and around standing and erupting filaments in both active regions and quiet regions suggest this picture (Moore and LaBonte, 1980; Hagyard et al., 1984a). SAMEX is designed to test this picture with greatly improved spatial and temporal resolution of the field structure in the photosphere, chromosphere, and corona, and to show how the sheared magnetic field configuration in and around a filament is formed over the course of several days.

velocitygrams is specifically suited to clarify these phenomena. The SAMEX observations of filaments and their surroundings will address basic questions such as:

- o Does the field configuration in and around quiet-region filaments have a different topology than that of active-region filaments?
- o Are quiet-region filament eruptions that produce no detectable chromospheric emission still basically flares? That is, are they just slow versions of mass-ejection flares in active regions?
- o Is flux cancellation at the photospheric neutral line (Martin et al., 1985) important in the process of forming the field configuration of quiet-region filaments? Are reconnection and flux submergence elements of this process? Does flux cancellation or reconnection unleash the field to drive the filament-field eruption as in Figure 6?

To carry out studies of filaments, both in quiet and active regions on the Sun, in addressing these questions we will need observations with all three SAMEX instruments over the whole filament. Typical lengths of prominences are on the order of 5 arc min (Allen, 1973); this then sets a requirement on the field of view of these instruments. The proposed field of the SAMEX magnetograph, 4.3' x 8.5', easily meets this requirement.

Attempts to measure the transverse field in the photosphere underneath quiet-region prominences with the MSFC vector magnetograph have been unsuccessful so far, primarily because of the low sensitivity (about 10^{-3}) of the system. This indicates that the sensitivity of the SAMEX instrument will have to be as high as present technology permits - the 10^{-4} that we propose - to attempt observations of the transverse field under quiet-region filaments.

3.3.3 Fine-scale magnetic structure and activity

In addition to large flares, the Sun exhibits a broad range of smaller-scale, flare-like activity. Transient fine-scale brightenings in chromospheric, transition-region, and coronal emissions have been observed in both active regions and the quiet Sun. The energy released (less than 10^{27} ergs) in any one of these events is far below that in even a small flare (10^{30} ergs), but taken together, these microflares may ultimately be found to account for the bulk of coronal heating. The extent of this activity has only recently become appreciated (Lin et al., 1981; Schadee et al., 1983; Porter et al., 1984, 1985), and our understanding of these events is quite incomplete. The scope, sensitivity, and resolution of the planned SAMEX instruments will allow a much better examination of these events and their magnetic origins.

Over most of the Sun's surface, the magnetic field in the photosphere and chromosphere is concentrated in a loose network outlining cells roughly 30,000 km in diameter. This magnetic network is formed by the horizontal outflow from the centers of supergranulation convection cells in and below the photosphere; the flow sweeps magnetic flux from the interiors to the boundaries of these convection cells. The network is present wherever the average field strength over areas larger than supergranules is less than about 100 gauss, which is everywhere outside of active regions that have sunspots. Magnetograms having resolution of a few arc seconds or better show that the photospheric field is strongly inhomogeneous within the lanes of this network (Dowdy et al., 1986). Most of the field is in intermittent clumps ranging in diameter from a few thousand kilometers down to less than a few hundred kilometers. Both polarities of flux are present and intermixed; some flux clumps appear to be unipolar, others show mixed polarity. Images of the chromosphere and transition region with arc second resolution show that the network above the photosphere is also strongly structured on the scale of the photospheric magnetic clumps: the network is seen to be full of loops and spicules ranging in size from about 10,000 km down to the limit of resolution. Spicules are strongly dynamic and transient, having velocities of tens of kilometers per second and lifetimes of minutes. A basic unsettled question is whether the magnetic activity that generates spicules also accomplishes the heating of the chromosphere, transition region, and corona. SAMEX will have sufficient magnetic sensitivity, spatial resolution, and time resolution of the photospheric magnetic field and of the chromosphere and corona to see if coronal heating is maintained by microflares associated with spicules in the magnetic network.

Time-sequenced magnetograms show that some of the flux in quiet regions is supplied locally by the emergence of tiny bipoles. These bipoles are called ephemeral active regions. They range in extent from as large as a supergranule to smaller than a few thousand kilometers, and live for less than a day. Ephemeral regions appear to be basically tiny versions of normal active regions. They produce enhanced heating of the corona (x-ray bright points), microflares, and macrospicules (Moore et al., 1977). This encourages the idea that the heating of the corona arises from fine-scale flare activity. In addition to microflares in ephemeral active regions, similar events in transition-region UV line emission occur at many smaller bipoles throughout the magnetic network (Porter et al., 1985, 1986, 1987). Likewise, microflares in active regions occur on localized magnetic inversion lines (Porter et al., 1984). That microflares in and out of active regions are seated on tiny inversion lines and emit x-rays and EUV radiation suggests that they are powered by dissipation of stored magnetic energy in a manner similar to that in larger flares. Study of the magnetic origins of these smaller members of the flare family may tell us whether the flare energy buildup and release mechanisms depend on the overall scale of the stressed magnetic field. This would be

basic new information on the physics of flares. The planned SAMEX instruments will show the fine-scale three-dimensional form, evolution, and action of the magnetic field from the photosphere to the corona with the degree of sensitivity and resolution necessary for this study. Neither the required coronal images nor the required seeing-free sub-arc second observations of the photosphere and chromosphere can be obtained from the ground.

3.3.4 Global aspects

Although both polarities of magnetic flux are mixed together everywhere in quiet regions, often one polarity dominates throughout an area a few hundred thousand kilometers across. In magnetograms with resolution worse than $10''$, these regions appear to be unipolar. The Sun's polar caps (latitude greater than 60°) are unipolar in this sense over most of each sunspot cycle, with opposite dominant polarity in northern and southern caps. The polarity of each cap reverses each cycle around the time of sunspot maximum. Other large unipolar areas come and go at lower latitudes with lifetimes of many solar rotations. Series of full-disk magnetograms spanning a solar cycle have shown that the low-latitude unipolar areas form from the remnants of active regions, and that the unipolar flux in the polar caps is built up by poleward migration of such low-latitude unipolar patches (Howard and LaBonte, 1981; Topka et al., 1982; Webb et al., 1984). This suggests that the large unipolar features are an important link in the dynamo process that sustains the solar cycle.

Most large unipolar areas are the seat of coronal holes, the source of high-speed streams in the solar wind. Thus, the large unipolar areas are important for the global structure of the corona and solar wind as well as for the solar cycle. It is therefore important to have long-term uninterrupted series of magnetograms from space to study the birth and evolution of the unipolar areas. How unipolar areas die, i.e., the process by which the unbalanced flux in unipolar areas is removed from the surface of the Sun, is presently unknown, probably because of the limitations on resolution and time coverage of ground-based observations.

Flux removal constitutes half of the solar cycle: over the course of each cycle, all of the flux supplied by emergence in active regions and ephemeral regions is also somehow removed. It seems necessary that most flux is removed by submergence, but this is yet to be demonstrated observationally (Moore and Rabin, 1985). Cancellation of opposite polarity flux by reconnection and submergence is a reasonable possibility for the dominant mode of flux removal both in active regions and in quiet regions. Such cancellation of the fine-scale clumps of opposite polarity in quiet regions might be the origin of the fine-scale magnetic activity that generates spicules and microflares and may heat the corona. To observe flux removal in the quiet network and

determine its relation to fine-scale magnetic activity in the chromosphere and corona requires (1) time-lapse magnetograms with uniform sub-arc second spatial resolution and time resolution less than the time scale of spicules, i.e., less than 5-10 minutes, and (2) chromospheric and coronal images with sub-arc second spatial resolution and a cadence of at least a few per minute to see what happens in microflares and the time development of spicules. To observe the details of flux removal in active regions also requires a vector magnetograph that provides high magnetic sensitivity and seeing-free sub-arc second observations. These observations can only be achieved with a spaceborne instrument complex such as that planned for SAMEX.

3.4 Predicting Solar Activity

Another dividend from the SAMEX mission will be the development of improved techniques for predicting solar activity as a result of the better understanding of the flare process that will come from the mission. The present state-of-the-art in solar forecasting can be characterized as an empirical-statistical procedure that lays considerable emphasis on forecaster experience in recognizing circumstances that are known to have been previously associated with flare activity of a given type and frequency (Sawyer et al., 1986). Magnetic configuration, sunspot group classification, rate of evolution, and history of activity in the active region during the previous day are among the parameters of greatest significance in forming a decision on the probable level of activity for the following 24 hours. Formal statistical methods for conditioning the data and deriving objective, quantitative forecasts have been attempted in various studies (see Smith and Neidig, 1986, and associated papers) with results that show little, if any, improvement over subjective techniques.

None of the forecasting methods in use at the present time are based on rigorous models of the active region magnetic field; this is as expected, in view of the lack both of a mature flare theory and of high quality data on which to apply the theory. Presently, synoptic data are acquired from numerous observatories, with the result that there is little uniformity in type, quality, and format of the data. In an effort to overcome this difficulty, the USAF Solar Electro-Optical Network (SEON) obtains synoptic data from five identical observatories spaced at longitudes appropriate for continuous solar monitoring. Still, nonuniformities are introduced by atmospheric conditions.

The collective effect of these observational deficiencies is to introduce a large amount of "system noise," with the result that observable changes in active-region magnetic fields are often statistically significant only on rather long time scales -- perhaps as much as a day in some cases. Theoretical studies (e.g., Wu et al., 1984), on the other hand, indicate that free energy in the active-region magnetic field can increase significantly on a time scale of less than 1 hour. Under these

circumstances the flare phenomenon might be expected to display an apparent stochasticism that would impose limitations on the accuracy of forecasts based on data of the present quality; this was the view expressed by Simon et al. (1979) and Stahl (1983). An obvious remedy for this situation is to obtain continuous, high-resolution data from space.

The new generation of data made possible by SAMEX will enable a serious effort at comparing time-dependent models of active-region magnetic fields with detailed observations of the photospheric vector field, as well as with chromospheric and coronal structures that trace out the fields at higher levels in the atmosphere. In this way calculations of currents in the force-free fields can be fine tuned until a best fit can be achieved with all the data. This procedure could be repeated on data sets collected hourly, for example, and would allow the field evolution to be tracked in time. In principle, the configuration could be extrapolated forward in time to determine if the trend is toward instability. Situations corresponding to colliding flux tubes and emerging or submerging flux could be treated similarly.

3.5 Summary

In this section on scientific objectives, we have presented arguments showing that if we are to address two very basic scientific objectives - understanding the buildup and storage of energy in the solar magnetic field and identifying the conditions necessary for the release of this energy in flares - we must observe, measure, and monitor the solar vector magnetic field in the photosphere with a space-based vector magnetograph.

We have argued that the specific objectives of this instrument should be to observe the formation and configuration of non-potential fields in photospheric active regions, to quantify the non-potential characteristics of these fields (e.g., angular shear, electric currents) in flaring and non-flaring regions, to quantify the free magnetic energy and the force-free nature of active-region fields as they develop and evolve, and to determine what factors lead to the destabilization of the fields and the eruption of flares.

We have seen that these objectives impose certain requirements on the vector magnetograph instrument: a temporal resolution of a few minutes, a field of view the size of typical active regions, spatial resolution of at least $0.5''$, polarimetric sensitivity on the order of 10^{-4} , and sufficient spectral resolution and spectral coverage to achieve this sensitivity, to measure line-of-sight velocities, and to perform flux tube diagnostics.

We have used these requirements as guidelines to prescribe the specific values for the instrumental parameters of the SAMEX vector magnetograph. Obviously we would like to fly an instrument

with the largest possible field of view, the highest possible temporal and spatial resolutions, full wavelength coverage over the visible solar spectrum with good spectral resolution, and the best possible magnetic sensitivity. However, limitations of present technology and the realities of space-based observations from a free-flying satellite impose certain restrictions on these desired specifications. Consequently the final parameters were derived from an extensive trade-off study that we made. We feel, however, that we have made an optimum compromise in our specification of these design parameters and that our specifications will allow us to carry out the primary scientific objectives of SAMEX better than any other instrumental design. In Table 5 we summarize these design parameters.

Table 5. Instrumental specifications of the vector magnetograph.

Spatial resolution	0.5 arc sec
Field of view	4.3 X 8.5 arc min
Full Sun imaging	through mapping (~ 32 images)
Polarization sensitivity	10^{-4}
Magnetic sensitivity	
longitudinal field	1 G in 38 sec
transverse field	< 50 G in 252 sec (~ 4 min)
Spectral parameters	
position accuracy	1 mA
FWHM	120 mA
range	5243.5 - 5254.0 A
Velocity resolution	60 m/sec
Temporal resolution	
vector magnetogram	~ 5 min ($S/N = 10^4$)
Dopplergram	10 sec

The field of view chosen (4.3' x 8.5') meets the primary condition of being large enough to completely cover the magnetic fields in most active regions. Figure 7 shows this field of view superposed on different areas of a Kitt Peak full-disk line-of-sight magnetogram and illustrates that this field of view is indeed quite adequate to overlay the regions shown. To investigate further the adequacy of the chosen field of view, we used results from a statistical study of the areas of active



Figure 7. The field of view of the SAMEX magnetograph superposed on a line-of-sight (longitudinal) magnetogram and H-alpha filtergram. The three separate fields of view shown cover two active regions and a large filament that were on the Sun on September 23, 1980, as seen in a Kitt Peak longitudinal magnetogram in the image on the left and a SOON (Solar Optical Observing Network) H-alpha filtergram on the right. These images of the typical Sun during maximum solar activity show why we need a field of view of 4 x 8 arc minutes: only a field this large will show the magnetic structure over the entire region and thus enable us to see simultaneous magnetic changes occurring in the region. According to a study by Tang et al. (1984), this field of view covers 90% of all bipolar regions.

regions covering a period of 14 years carried out by Tang et al. (1984). They derived a formula for the number of bipolar regions $N(A)$ with areas A greater than or equal to the area A (in square arc minutes) contained within the 10-G contours of the Mt. Wilson line-of-sight magnetograms; this formula can be written

$$N(A) = N_0 e^{-A/13.63},$$

where N_0 is the total number of bipolar regions observed. We used this formula to determine the percent of bipolar regions that would be contained within different fields of view for 1:1 and 2:1 aspect ratios; these results are given in Table 6. These data show that 90% of all bipolar regions would be contained within the SAMEX field of view.

Table 6. Percent of active regions contained within different fields of view.

Field of view		Percent of active regions contained within the field of view
Aspect ratio = 1:1	Aspect ratio = 2:1	
2.8' x 2.8'	4.0' x 2.0'	44.4
3.1' x 3.1'	4.3' x 2.2'	50.0
3.7' x 3.7'	5.2' x 2.6'	63.2
5.6' x 5.6'	7.9' x 4.0'	90.0
6.0' x 6.0'	8.5' x 4.2'	93.0

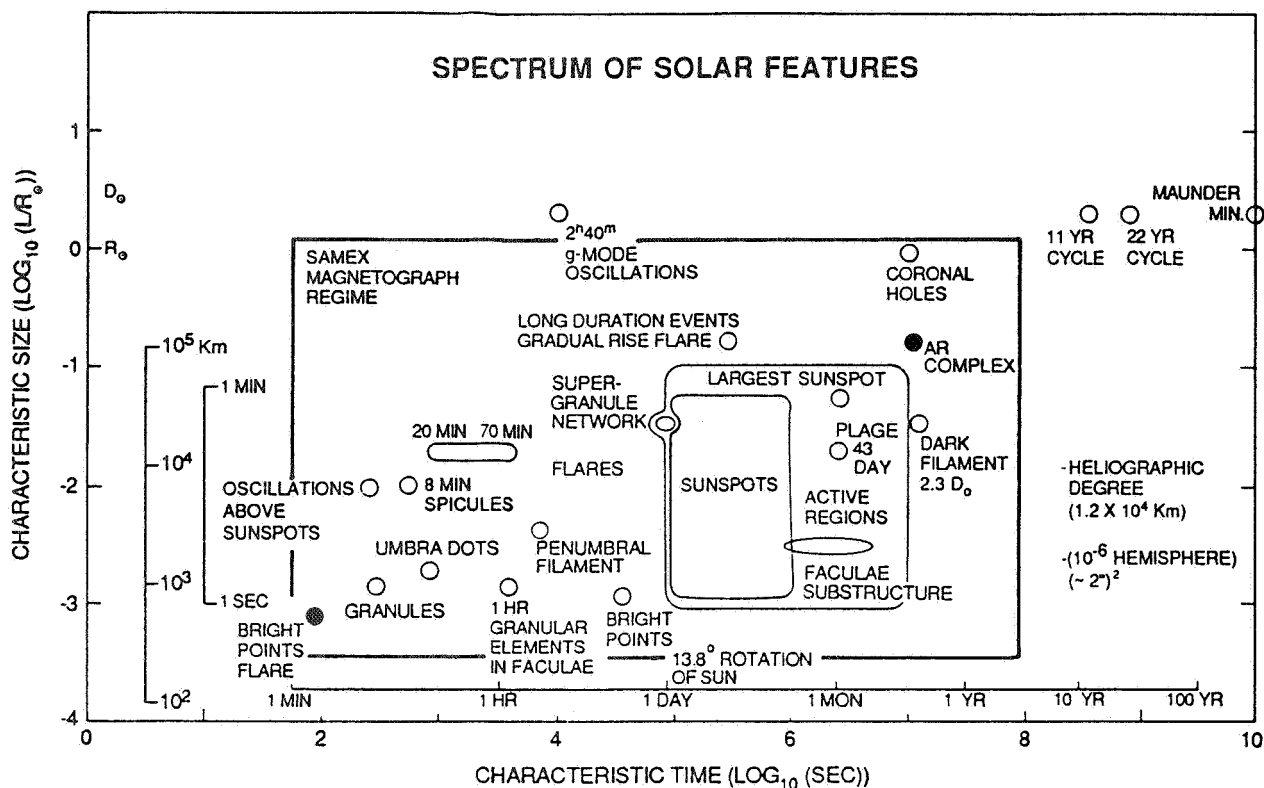
Full coverage of an active region is absolutely necessary for two very important reasons. First, while the presence of angular shear will indicate areas with flare potential, we will want to study the fields away from sites of flare onset to look for any flare-related changes in the fields. Also, we have indicated previously the problems encountered in model calculations if the boundary conditions of the model are not met because the field in the active region is not measured in its entirety. The areal coverage of our field of view is also optimum for mapping the full solar disk by overlapping images (it will take about 32 images to map the Sun). This full-disk mapping would be more time consuming for a smaller field of view since many more images would be needed.

We have chosen a spatial resolution of 0.5" for the instrument. While this resolution will not resolve the finest strands in the network magnetic field that are composed of 1-2 kG fields in 0.1-0.4 arc sec elements (Stenflo, 1973), it will resolve most features observed in active regions where our primary interest will be centered. In Table 7 we list various solar features and give their spatial dimensions and approximate lifetimes; these data are summarized graphically in Figure 8.

Table 7. Sizes and lifetimes of solar features.

Feature	Size	Lifetime
Superpenumbral fibrils	0.3" - 3.0"	< 20 min
Dark penumbral striations	~ 0.2"	
Penumbral bright grains	≤ 0.4"	60 min
Umbral dots	0.4" - 0.5"	60 min
Magnetic sunspot features	≤ 0.5"	
Pore	2" - 5"	10 - 15 min
Magnetic knot	~ 1.4"	60 min
Filagree	≤ 0.4"	5 - 20 min
Fibrils	0.9" - 2.8"	3 - 50 min
Arch filament system	1.4" - 4.0"	30 min
Structure of flare kernel	≤ 0.2"	
Granulation	1" - 2"	5 - 20 min
Filagree	≤ 0.4"	5 - 20 min
Supergranule	40" - 50"	~ 1 day
Network elements	0.2" - 2"	5 min - 2 hr
Spicules	0.7" - 1.6"	5 - 10 min
Internal network field	~ 0.5"	a few hours
Ephemeral regions	a few arc sec	~ 1 day

There are very good reasons relevant to the SAMEX scientific objectives that argue against compromising the 0.5" spatial resolution by going to a smaller (and less expensive) telescope. (1) To calculate the free magnetic energy as accurately as possible we need to resolve the ~0.5 arc sec "salt and pepper" field. (2) We want to resolve the smallest elements of emerging flux to determine whether they are possibly involved in triggering the flare process. (3) As we come closer to



resolving flux elements we improve the accuracy of measuring the field. (4) We would like the spatial resolution to be about equal to the scale height of formation of the spectral lines used in measuring the field. For the Fe I 5250.22 line this is about 0.5" according to calculations we have performed for contribution functions for this line. (5) We know that the very best ground-based observations show magnetic structure on a scale of 0.5" (Figure 9); we obviously want to do this well or better from space where we can observe with the 0.5" resolution over long periods of time, much longer than we could ever expect to achieve from the ground.

We determined that a time resolution of ~5 minutes is needed to study the dynamical aspects of solar activity and the evolution of the magnetic features whose lifetimes are listed in Table 7. The data in this table show that the time span of 5 minutes is below the lifetimes of most of these features. However, preflare impulsive phenomena can occur on the order of 10 minutes prior to flares, and rise times seen in soft x-ray data generally fall in the range of 1 to 5 minutes. Thus a temporal resolution of 5 minutes or better is needed to observe these important aspects of the preflare state and flare onset. Other dynamical features of solar activity requiring temporal resolutions better than 5 minutes include penumbral waves with periods of about 4 minutes and umbral oscillations whose periods range from 1 to 8 minutes. To observe these and other phenomena with periods/lifetimes less than 5 minutes will mean that we will have to sacrifice magnetic sensitivity to attain better than a 5-minute time resolution. The temporal resolution of ~5 minutes will also prevent image smearing due to solar rotation and insure that phenomena traveling horizontally with speeds below the acoustic velocity of a few km per second will be resolved.

We have opted for a limited spectral range for our observations of photospheric magnetic fields. Specifically we have selected the wavelength interval from 5243.5 to 5254.0 Å, a range of only 10.5 Å. However, this spectral range includes the three lines selected by Stenflo (private communication) as the best lines in the visible spectrum for performing flux tube diagnostics. There are many good reasons for limiting the spectral range of the instrument. (1) The problem of thermal control of the instrumentation is greatly reduced if we include a full-aperture prefilter that rejects all wavelengths of incident light except those in this narrow interval. This degree of heat rejection would not be possible for an instrument designed to admit all visible wavelengths. (2) This narrow spectral range minimizes the design problems for achromatic waveplates needed in the polarization analysis. (3) It is imperative that we achieve an extremely low level of instrumental polarization to measure the transverse magnetic field with the accuracy we propose. This low level can only be achieved by coating the lenses and mirrors of the optics, and these coatings are effective over only a small wavelength interval. (4) If the filter selected for the instrument is a birefringent filter, this small spectral range will

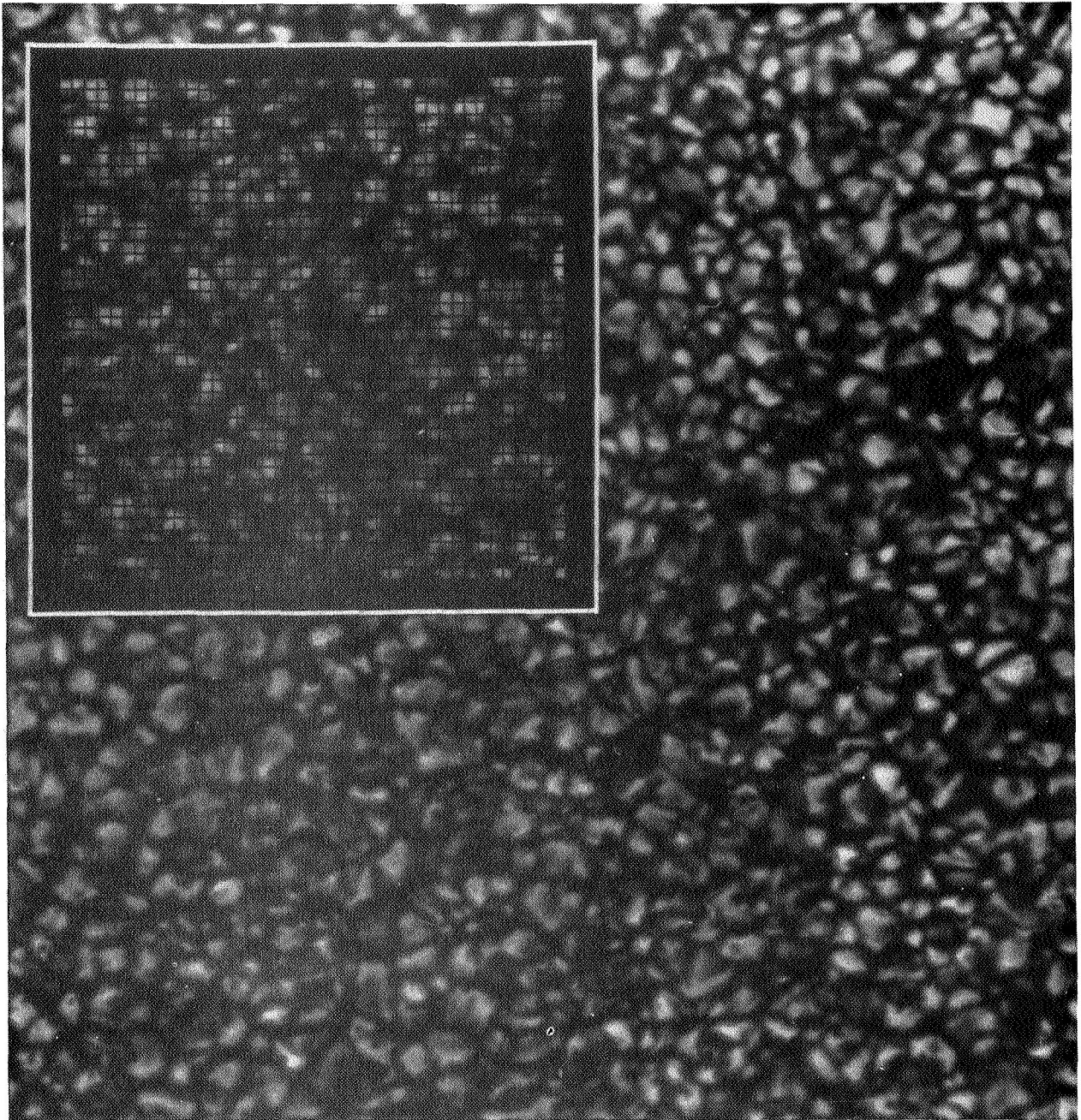


Figure 9. The SAMEX spatial resolution. The photospheric granulation pattern at sub-quarter-arc sec resolution is shown with a 10 x 10 arc sec grid of quarter-arc sec divisions superimposed. The grid is representative of a section of the total 256 x 512 arc sec pixel array for the SAMEX detector. The granulation image of 1-2 arc sec cells was obtained under ideal atmospheric conditions with the Vacuum Solar Telescope of the Royal Swedish Academy of Sciences at La Palma, Canary Islands. The SAMEX magnetograph will be able to resolve the magnetic structure of these granulation cells by providing stable images over the period of time necessary to acquire a vector magnetogram.

reduce the overall length of the filter and result in a very simple design. (5) If a Fabry-Perot filter is selected, the narrow range will minimize requirements on the necessary prefilters.

We have specified that the spectral bandpass be on the order of 120 mÅ. Narrower bandpasses can be achieved: a filter with a bandpass of 60 mÅ was flown on Spacelab 2 (Title et al., 1986). However we believe the broader bandpass is better suited for the SAMEX mission. Calculations show that it will give a magnetic sensitivity comparable to that of the narrower filter (see section II.5), it will transmit more light to the detector, and there is more room for error in centering a broader filter at the wavelength in the spectral line where the slope is steepest and where the magnetic signal is greatest (this is discussed in section II.5 also).

Finally, we are convinced that the most important requirement for the SAMEX vector magnetograph is that it measure the transverse component of the magnetic field with the highest accuracy possible, and we have determined that a polarimetric sensitivity of 10^{-4} , which is an order of magnitude better than that of any present instrument, can be achieved. Throughout our discussion of the SAMEX scientific objectives we have presented good, solid arguments for achieving this level of accuracy in measuring the vector magnetic field: (1) This level of sensitivity will permit measurements of transverse fields that are much weaker than those measured with present-day, conventional instruments. (2) There will be a significant improvement in measurements of the azimuth of the field, the parameter that is the primary signal of stressed, non-potential fields. (3) We will reduce significantly the errors in calculating electric currents, energy, Lorentz forces and torques, and gradients. (4) Model calculations will be more accurate. (5) Observations of the transverse field in emerging and submerging flux and in flux cancellation will be possible for the first time. (6) Detection of changes in the magnetic field as a result of flares will be possible for more flares.

In conclusion, we believe that an instrument meeting these specifications on magnetic sensitivity, field of view, spectral coverage, and spatial and temporal resolutions will achieve the scientific objectives of the SAMEX mission and will produce many exciting new discoveries about the Sun. The SAMEX magnetograph will be the first vector magnetograph flown successfully in space, and it may be the only one flown in this century. For that reason alone, we believe that it should be the very best possible instrument that present technology can produce.

II. THE VECTOR MAGNETOGRAPH INSTRUMENT

1. Introduction

In this section we will describe the design we have developed for the SAMEX vector magnetograph based on the instrumental specifications listed in Table 5. The design that is presented covers the imaging optics, polarimeter, spectral filter, and detector. This section concludes with a summary of the overall performance that we expect from this particular design. A large portion of the details of this design is based upon the experience gained by the MSFC Solar Science Group in testing, calibrating, and operating the Marshall Space Flight Center vector magnetograph which has been in operation for over a decade.

1.1 Specifications of the Overall Design

The instrumental specifications in Table 5 have determined the basic design of the magnetograph system since, to measure the vector magnetic field over an entire active region with the desired temporal resolution of 5 minutes and with 0.5" spatial resolution, a filter-type magnetograph is preferred over a Stokes polarimeter. The basic operational principle of a filter magnetograph is an analysis of the states of polarization of Zeeman-sensitive absorption lines at selected wavelengths across the spectral lines. When formed in a magnetic field, these lines are split into several polarized components; the amount of splitting is proportional to the field intensity, and each component is characterized by a specific state of linear, circular or elliptical polarization depending on whether the magnetic field is transverse to, along, or inclined to the line-of-sight, respectively. In most cases the intensity of the magnetic field is insufficient to completely separate the polarized components, so the Zeeman splitting cannot be resolved and measured with a filter magnetograph. However, the net linearly and circularly polarized intensities, measured at some wavelength in the spectral line, can be quantitatively related to the intensities of the transverse and line-of-sight (longitudinal) field components, respectively. This comes about because, as the field strength increases, the different polarized Zeeman components become more and more unblended with a resulting increase of measured polarized intensity at a given wavelength. The quantitative relations relating the measured net polarized intensities to the longitudinal and transverse field components are derived by solving the equations describing the radiative transfer of polarized light through the solar atmosphere. Four quantities are required to completely specify the intensity and polarization of a beam of light. A commonly used representation is the Stokes vector whose four components - I, Q, U, and V - are measures of the total intensity (I), two linearly polarized intensities (Q and U) polarized at 45° relative to each other, and the circularly polarized intensity (V). Four coupled

radiative transfer equations can be developed for these four Stokes parameters (Unno, 1956; Beckers, 1969; Landi Degl'Innocenti, 1983). Expressions for the emergent intensities I, Q, U, and V as functions of the magnetic field vector are then obtained by solving the coupled transfer equations under certain simplifying assumptions. These functional relations between Stokes intensities and magnetic field are then used to derive the longitudinal and transverse components of the magnetic vector from measurements of net linearly and circularly polarized intensities. The third component of the vector field, the azimuth of the transverse component, is obtained directly from a determination of the plane of linear polarization of the spectral intensity. An exposition of these analytical techniques is presented in Appendix D.

Since a filter magnetograph measures linear and circular polarizations, a basic component of a filter system is a polarimeter that accurately determines the I, Q, U, and V intensities. Indeed, the accuracy to which the polarimeter measures these Stokes parameters will determine the ultimate magnetic sensitivity of the flight instrument. This relation between polarimetric sensitivity and the noise level in magnetic field measurements is seen in Figure 10 which shows the variation of magnetic field strength with the observed degree of polarization for a photospheric and a penumbral model. This graph indicates that, to measure the longitudinal field to less than 1 gauss and the transverse field to a few tens of gauss, the degree of polarization will have to be measured to 10^{-4} or better. This means that the instrumental polarization of the optical system will have to be reduced to a figure of 10^{-5} or less. This figure is the design specification we have established for the optical system of the SAMEX magnetograph.

Magnetic sensitivity is also determined by the basic signal-to-noise ratio for the system; this is generally governed by the detector. The filter magnetograph requires a two-dimensional detector to image the polarized intensities over the field of view of an active region. Detectors most suited for this purpose are large-array ($\sim 2000 \times 2000$ pixels) solid-state devices that have signal-to-noise ratios commonly of the order 400-600. Increasing this ratio to 10^4 so that weak transverse magnetic fields can be measured with low uncertainty requires adding successive polarization images (i.e., a process of image enhancement). The limitation on how many images are added together is governed by the temporal resolution desired, and this depends on the exposure time required for the detector array (e.g., a CCD chip) and on the time required to read out the array. Estimations based on quoted CCD properties and on assumed optical transmissions of the optical components indicate that the CCD exposure time should be about 0.37 sec. On the other hand, to read out a 1024×2048 array of pixels at a rate of ~ 1 microsec per pixel would take 2.1 sec, but this time can be reduced below the exposure time of 0.37 sec by using multiple readout ports serviced with individual A/D converters. With at

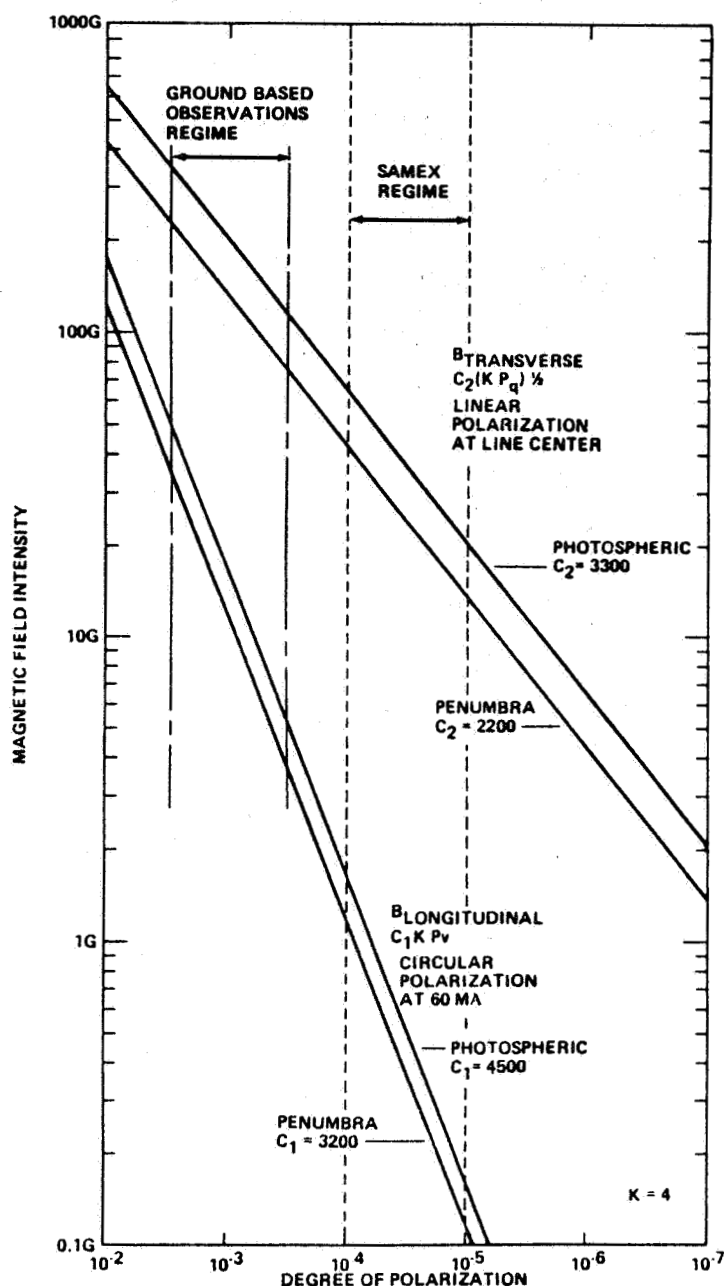


Figure 10. The photospheric longitudinal and transverse magnetic field strength derived from the measured degree of circular and linear polarization, respectively, for quiet photosphere and penumbral models. The parameters C_1 and C_2 are specified by the assumed model for the solar atmosphere and the radiative transfer through it, for the case of the Fe I 5250.2 line with $g = 3$ (cf. Hagyard and West, 1982, and Appendix D). The parameter k calibrates the magnetograph measurements with sunspot fields derived from spectrographic observations. The transverse curves are for the spectral filter tuned to a position at line center, and the line-of-sight curves are for the filter at 60 mÅ in the wing of the line. The regime of minimum detectable polarization is shown for ground-based observations and can be compared with that for the proposed SAMEX instrument.

least four to six such parallel readout ports, the readout time for a CCD-type detector is estimated at 0.35 sec. Since the image exposure and readout will be done in parallel, the larger of the times for these two processes, 0.37 sec, will be the factor that determines the temporal resolution. Within 5 minutes, both circular images can be enhanced 50 times and the three linear images 225 times. The resulting complete vector magnetogram would have a 1 G sensitivity in the line-of-sight component and a 35-50 G sensitivity in the transverse component, with a temporal resolution of 5 minutes which is the temporal resolution required by the scientific objectives. To obtain better time resolution by reducing the number of images added together would result in less magnetic sensitivity. Conversely, better magnetic sensitivity would accordingly require a loss of temporal resolution. Based on these considerations, we believe that vector magnetograms with longitudinal and transverse field sensitivities of 1 and 35-50 G, respectively, can be obtained for active regions in 5 minutes with a filter magnetograph that has an instrumental polarization less than 10^{-5} and that can achieve a signal-to-noise ratio of 10^4 in the linear measurements through the superposition of approximately 225 images. These levels of magnetic sensitivity are within the requirements of our scientific objectives and represent a significant improvement over present ground-based measurements where noise levels in the transverse fields are typically 125-180 G, a limitation imposed by instrumental polarization in the optics and polarimeter.

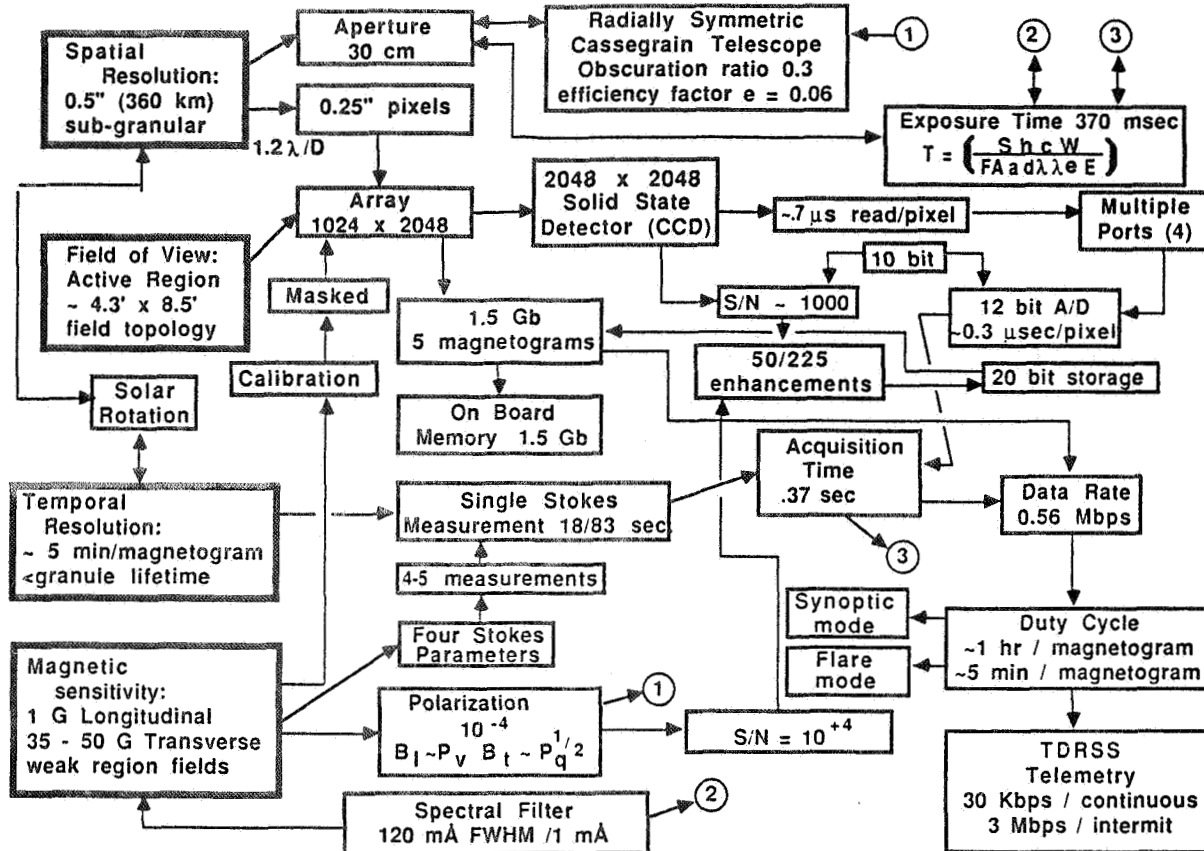
The required spatial resolution of 0.5" is easily obtained with a moderate-size (30 cm) telescope. However, there is a trade-off between the spatial resolution achieved and the field of view covered, given the present limitations on the number of resolution elements in two-dimensional detector arrays. With a detector having a 2048 x 2048 pixel array (or a matrix of four detector arrays each composed of 1024 x 1024 pixels) where half of the array is used for imaging and half for temporary storage of the data from the imaging section, a field of view of 4.3 x 8.5 arc min can be imaged with 0.25" pixel resolution and 0.5" spatial resolution. To achieve a larger field of view with the same detector array would require a proportionate loss in spatial resolution which is not desirable. A larger field with 0.5" spatial resolution could be realized by combining more individual arrays in a matrix but this would impact onboard data storage requirements. For a 1024 x 2048 array, the total number of imaged pixels would be 2.1×10^6 with approximately 20 bits of information for each pixel. Each vector magnetogram would consist of four such arrays and, since it would take about 5 minutes to acquire the data, the resulting data rate would be 0.56 Mbps for just the SAMEX magnetograph; the H-alpha telescope and x-ray imager will also add to the overall data rate. The NASA Tracking Data and Relay Satellite System (TDRSS) allows a nominal data rate of 50 Kbps on its multiple-access S-band channel. It could not accommodate data rates on the order of 0.5 Mbps and onboard data storage would be necessary. Using larger arrays would produce even larger data rates and require much

larger data storage capabilities. Based on these considerations, we have determined that the 2048 x 2048 array with 0.5" spatial resolution achieves an optimum trade-off among spatial resolution, areal coverage, and experiment cadence, given the limitations imposed by available technology. These general design relationships are shown in the form of a flow chart in Figure 11.

With this selection of field of view and spatial resolution, the SAMEX magnetograph represents a significant improvement over present instruments and a major advancement in operations by providing almost continuous observational coverage of the Sun. The best available vector magnetograms taken from the ground have spatial resolutions at about the 2" level. The Earth's atmosphere limits the ultimate ground-based resolution to about 1" except on rare occasions when excellent seeing permits about 0.3" resolution. However, because each vector magnetogram requires the measurement of at least four polarized intensity images, and a number of each of these images must be added together to attain sufficient magnetic sensitivity in the transverse component, the differences in seeing that would occur between observations of successive polarization states would make anything below 1" resolution extremely difficult to achieve. With a constant spatial resolution of 0.5" over the time interval of a single observation, the SAMEX filter magnetograph will be four times better at measuring the smallest flux elements than any ground-based systems, either planned or presently in operation. Other systems cannot match SAMEX's spatial resolution over the course of a single observation. In addition, the SAMEX system will measure the transverse field with a magnetic sensitivity of ~35 G which is 3.5 times better than the performance of existing vector magnetographs (~125 G sensitivity). Considering both the improved magnetic sensitivity and spatial resolution of the SAMEX instrument, the measurement of magnetic fields will be improved over ground-based measurements by a factor of 14! Thus the SAMEX instrument represents a significant advance in measuring the vector fields of active regions with high spatial resolution over long intervals of time. With no effects of atmospheric distortion, changes observed in magnetograms taken at different times can be attributed to real changes in the solar field.

While a 0.5" (360 km) spatial resolution prohibits the resolution of the finest strands in the magnetic field that are composed of 1-2 kG fields in 100-300 km elements (Stenflo, 1973), we will be able to identify the filling factor of these fine-scale fields by observing in several different spectral lines and thus estimate the scale sizes of the magnetic structures below the limit imposed by the 0.5" spatial resolution of the instrument. This method requires a filter that is tunable over a spectral range that includes the necessary lines for such an analysis. The three Fe I lines at 5247, 5250.2, and 5250.6 Å are ideal for this analysis (Stenflo, private communication) and

DESIGN RELATIONSHIPS



5-161-6-2C

Figure 11. The overall interrelations of the SAMEX magnetograph requirements. This diagram shows how the design specifications for spatial resolution, field of view, temporal resolution, and magnetic sensitivity are interrelated with each other and produce requirements on the magnetograph instrument, onboard computers, and spacecraft systems. To satisfy most of these requirements in an optimum sense requires many different trade-offs. The details of these relationships are described in the text.

ORIGINAL PAGE IS
OF POOR QUALITY

their mutual proximity imposes only minimal requirements on the spectral range of the filter.

There is an important trade-off between the extent of the filter's spectral range and magnetic sensitivity which argues for a limited spectral range. As we indicated previously, to obtain the required level of magnetic sensitivity, the magnetograph's instrumental polarization will have to be reduced to a level of 10^{-5} or less. This can be accomplished by coating the mirrors and lenses that comprise the telescope and foreoptics of the magnetograph with special coatings. But the effectiveness of these coatings is a function of wavelength, so, if the filter is designed to cover an extended spectral range, there will be only a small fraction of this range where the coatings will have maximum effect. Consequently, if we are to achieve the most important objective of high magnetic sensitivity without resorting to a very complicated optical system, the spectral range of the filter where magnetic measurements are taken must be restricted to a relatively narrow portion of the visible spectrum.

Calculations of polarization signals as functions of spectral bandpass show that magnetic sensitivity increases as the bandpass decreases. However, as we show in section II.5, if the bandpass is too narrow, the polarization signal is a very sensitive function of wavelength, with the result that any slight offset in wavelength of the spectral filter produces a significant drop in signal. We believe that this behavior offsets the gain in magnetic sensitivity for very narrow bandpass filters, and we conclude that a spectral bandpass of 120 mÅ will provide the necessary sensitivity. This bandpass requirement also reduces the complexity of the design of the spectral filter significantly.

Many of our scientific objectives require measurements of line-of-sight velocities in the photosphere. These data can be obtained by incorporating a Doppler mode in the operating sequences of the vector magnetograph. In this mode intensities in the red and blue wings of a spectral line are measured by rapidly tuning the spectral filter to opposite wings of the line. If the spectral filter's wavelength position can be adjusted with an accuracy of 1 mÅ, the instrument should be able to resolve velocities of 60 m/sec.

1.2 Radiometry Considerations and Temporal Resolution

The total exposure time required to obtain one vector magnetogram is a basic parameter that we have used in determining many of the operational specifications of the SAMEX instrument. It depends on the characteristics of the detector system and the expected solar irradiance passing through the SAMEX optical system, i.e., the radiometry of the system. Here we calculate the exposure times for the SAMEX magnetograph, based on a system with a 30 cm aperture telescope and 0.3 obscuration ratio, a 4.3

x 8.5 arc minute field of view, and a 1024 x 2048 imaging CCD array. We have assumed typical (quoted) values for the CCD's quantum efficiency, well depth, noise values, and charge transfer times. In performing the radiometric calculations, we have used optimum but realistic values for the optical transmissions of the various lenses, filters and wave plates comprising the SAMEX optics. We have checked the calculations by applying them to the MSFC vector magnetograph.

The exposure time is determined essentially by the time it takes to fill each pixel well with electrons. This time is governed by the well capacity of the CCD and the number of electrons generated each second by the light incident on each pixel. This latter parameter is itself determined by the quantum efficiency (yield) of the CCD and the number of photons, n , incident each second on the active area of each pixel. To determine n , we consider the total radiant energy incident on the active area of the detector each second, P_i . P_i is equal to the solar radiant power per unit area per unit wavelength interval at the wavelength of interest, $P(\lambda)$, multiplied by the wavelength interval and the effective area, and corrected for transmission losses:

$$P_i = P(\lambda) \Delta\lambda A f(A) T .$$

In this equation, $P(\lambda)$ is the total solar irradiance at the wavelength λ , $\Delta\lambda$ is the wavelength interval of the filter's spectral bandpass (120 mÅ), A is the effective collecting area of the telescope, $f(A)$ gives the fraction of the collected area that is imaged at the detector's faceplate, and T represents the net optical transmission from the Sun to the CCD.

The incident solar irradiance in the continuum at the wavelength interval of interest (5244-5254 Å) is $2000 \text{ W m}^{-2} \mu\text{m}^{-1}$. However, magnetograph observations are taken in a Zeeman-sensitive absorption line. For the Fe I line at 5250.2 Å, the minimum irradiance at the center of the line is approximately $700 \text{ W m}^{-2} \mu\text{m}^{-1}$. If we assume the observations are made with the spectral filter centered near the region of steepest slope in the line profile, we get approximately $1400 \text{ W m}^{-2} \mu\text{m}^{-1}$ for $P(\lambda)$.

The collecting area of the telescope is given by

$$A = \pi(R_1^2 - R_2^2) ,$$

where R_1 is 15 cm and R_2 is $0.3 R_1$.

The fraction of the total area collected by the telescope that is imaged at the detector is equal to the ratio of the area of the field of view to the solar area:

$$f(A) = 4.3' \times 8.5' / (\pi \times 16' \times 16') = 0.0453 \quad .$$

The total optical transmission is the product of the transmissions of the individual elements in the optical path; these values are listed in Table 8. Values for the transmission of a birefringent filter were assumed for the filter. For a solid Fabry-Perot filter the transmission value could be increased (see section II.5); however, a somewhat conservative value is used here.

Table 8. Transmission Values for SAMEX Optics

Element	Transmission	
Full-aperture prefilter	0.90	
Telescope	0.90	
Relay optics	0.98	
Quarter-wave plate	0.95	
Rotating polarizer	0.48	
{ Order-sorting filter	0.80 }	Birefringent filter
{ Spectral filter	0.25 }	
Imaging optics	0.95	
CCD window	0.98	

The net transmission T from these values is 0.067.

Using these data, we have for P_i :

$$P_i = (1400 \text{ W m}^{-2} \mu\text{m}^{-1})(1.2 \times 10^{-5} \mu\text{m})(0.064324 \text{ m}^2)(0.0453)(0.067) \quad ,$$

or, there are 3.26×10^{-6} Joule of photon energy incident each second on the CCD faceplate. At a wavelength of 5250 Å, each photon has an energy of 3.78×10^{-19} Joule, so there is a flux of 8.62×10^{12} photons per second imaged over the CCD faceplate.

These data allow us to estimate n, since there are 1024 x 2048 pixels on the CCD faceplate that are exposed to this incident flux of photons. The active area of each pixel is $21 \times 27 \mu\text{m}^2$, although the total area occupied by each pixel is $27 \times 27 \mu\text{m}^2$. Hence, for the value of n we have the following:

$$n = (21/27)(8.62 \times 10^{12} \text{ photons/sec}) / (2.0972 \times 10^6 \text{ pixels}) \quad ,$$

so that there are 3.20×10^6 photons incident on each pixel per second.

The quoted quantum efficiency for the 2048 x 2048 Tektronics CCD is 50%, which gives 1.60×10^6 electrons entering a pixel well each second. The well capacity for this particular chip is given as 7.5×10^5 electrons. This means that the exposure time will be approximately 0.37 sec to fill the well to 80% of capacity. This is the basic value for the exposure time we will take for each polarized image.

The acquisition time for a complete vector magnetogram is determined by both the total number of polarized images needed for a magnetogram and the number of times each polarized image must be enhanced to reduce the inherent detector noise to a value comparable to the polarization accuracy desired, i.e., 10^{-4} . The intrinsic Poisson statistical noise of the CCD detector will limit the accuracy of a single polarization measurement. For a detector with a well depth of 7.5×10^5 electrons, the signal at 80% of this capacity is 6×10^5 electrons, so that the statistical noise will be $(6 \times 10^5)^{1/2}$. Thus, the ratio of signal to noise is 775, a value an order of magnitude lower than the required value of 10^4 . Moreover, a more detailed analysis of the noise from a CCD device reduces the signal-to-noise ratio by 12% to a value of 667 (see section II.6 on the detector for more details). This value for signal to noise indicates that the polarization accuracy for one image of polarized intensity will be on the order of 1.5×10^{-3} . To achieve a sensitivity of 10^{-4} , 225 such images will have to be added together to reduce the statistical noise sufficiently.

Since we want all the sensitivity we can get in the linear polarization measurements, we can expect that the linearly polarized data will have to be enhanced 225 times. However, in the case of the line-of-sight component of the magnetic field which is derived from the circularly polarized intensities, enhancing 225 times would be unnecessary - calculations indicate that the line-of-sight field could be measured to a 1-gauss accuracy with only 50 image enhancements.

These considerations indicate what the overall temporal resolution of the SAMEX magnetograph will be to acquire a single vector magnetogram. For a vector magnetogram, five polarized images are needed, two circular polarization measurements, and three linear:

$$I + V, \quad I - V, \quad I + Q, \quad I - Q, \quad \text{and} \quad I + U.$$

These data then provide the information to derive the three polarization intensities of the Stokes vector (Q , U , and V) and two intensities (I_V and I_Q). (The intensity I_V is obtained by the addition of $(I + V)$ and $(I - V)$ and similarly for I_Q .) Both intensities are necessary: I_Q is needed to derive Q and U from the linearly polarized intensities that are enhanced 225 times ($I \pm Q$, $I \pm U$); I_V is needed to derive V from the circularly polarized intensities ($I \pm V$) that are only enhanced 50 times. If the enhancement process was precisely linear, only one

intensity would be needed. However, because of statistical noise, the process is not entirely linear and thus we require both I_Q and I_V . This additional extra measurement represents only a 7% increase in the total time to obtain a vector magnetogram.

The timing for the vector magnetogram is thus determined by the time to expose the CCD chip for 0.37 sec per image, enhance the 2 circularly polarized images 50 times (each), enhance each of the 3 linear measurements 225 times, and read out each image from the CCD. Since we propose to use one half of the 2048 x 2048 array for imaging and to shift each image to the other one half of the array where it is read out while the next image is being exposed, the exposure and readout processes are done in parallel and the timing is determined by the longer of the two processes. If we assume the time to read each pixel is 1 microsec, then the time to shift 1024 pixels in 2048 parallel rows is ~ 0.001 sec. As the next image is being exposed for 0.37 sec, the shifted image is read out using four to six parallel A/D channels so that the total time to read the shifted image is 0.33 sec. This means that it is the exposure time that determines the temporal resolution, T:

$$T = (2 \times 50 + 3 \times 225) \times (0.001 + 0.37) \text{ sec} = 290 \text{ sec} ,$$

or about 5 minutes. (A reduction to a set of four measurements would eliminate only 50 enhancements since the I-V measurement would be discarded rather than the I-Q one because we want to calculate the intensities I, Q, and U from the set of data with the maximum S/N. Since discarding 50 images represents a saving of only 19 sec, the basic set of five measured intensities will be baselined throughout the report.)

This radiometric calculation indicates that we can achieve the desired accuracy in measurements of the vector magnetic field and still attain a reasonable temporal resolution.

1.3 Image Quality

The radiometric considerations discussed in the previous section were derived for an optical design using a 30-cm Cassegrain telescope for the collecting element of the magnetograph. The diffraction-limited resolution of this size telescope is approximately 0.5 arc sec which is achieved in the final image by using a detector pixel array with 0.25 arc sec size pixels. To investigate the image quality of this system and to lay a basis for the optical design analysis in the following section, we analyze this proposed optical design in this section through the use of transfer function theory which provides a measure of performance in terms of resolution. Specifically, the modulation transfer function (MTF) is used. The MTF is defined as the ratio of the brightness modulation in the image to that in the object as a function of spatial frequency (cycles per unit length) of a sine wave pattern. The usefulness of this theory

comes from the result that the system function is a product of the optical, detector, and electronics' MTF's. For the present analysis the telescope will be approximated by an annular aperture with negligible wavefront errors. The ratio of the secondary mirror diameter to the primary mirror diameter, the obscuration ratio, is 0.3. This value minimizes the angles of incidence in a Cassegrain telescope by making them approximately equal on the secondary and primary. The image quality for the vector magnetograph optical system using the modulation transfer function (MTF) is shown in Figure 12. In Figure 12a the MTF's are shown for a 30-cm circular aperture for reference as a dashed line and for 30-cm annular apertures with obscuration ratios of 0.2, 0.3, and 0.4 as solid lines. The MTF is the autocorrelation function of the pupil function which is equivalent to the geometric process of computing the area common to two similar apertures displaced in spatial frequency. The annular apertures are equivalent to an aberration-free Cassegrain telescope. The high frequency responses of the annular apertures are a little higher than an equivalent circular aperture as a result of the central obscuration. The effects of the obscuration ratio do not change the limiting resolution of the system which is 2.7 cycles per arc sec and is determined strictly by the aperture diameter. The conversion between cycles per arc sec and cycles per millimeter was based on the 27-micron detector pixel subtending an angle of 0.25 arc sec in object space.

In Figure 12b the MTF's are shown for 30-, 35-, and 40-cm annular apertures, for the same obscuration ratio of 0.3, as solid lines; the circular apertures for these diameters are shown as dashed lines for reference. The MTF for a 0.25 arc sec pixel detector array is shown as a dashed-dotted line. The MTF of a detector with a pixel size "a" is given by $\sin(\pi a u)/(\pi a u)$, where u is the spatial frequency. The MTF for a detector array was derived by White (1976). It is seen that the telescope is less responsive and limits the resolution of the system along with the sampling frequency of the detector. The detector MTF must be better than that of the telescope in order to reach the limiting resolution of the telescope since the total resolution of the system is the product of the MTF's. In the present configuration the aliasing due to the detector array does not occur due to low pass filtering by the telescope (Barbe and Campana, 1976).

In Figure 12c the MTF's are shown for a fixed f-ratio of 74.25 for annular apertures of 30, 35, and 40 cm with an obscuration ratio of 0.3. The f-ratio was fixed at the value determined by the 0.25 arc sec pixel with a physical dimension of 27 microns. Equivalent pixel sizes in arc sec (and focal lengths in meters) are 0.25 (22.27), 0.21 (25.99), and 0.19 (29.70), respectively, for the three aperture sizes and a 27-micron pixel. For a constant f-ratio the limiting resolution in cycles/mm is constant. However, as a function of cycles/arc sec, the limiting resolution varies since the relationship between the physical dimension of the pixel and the subtended angular

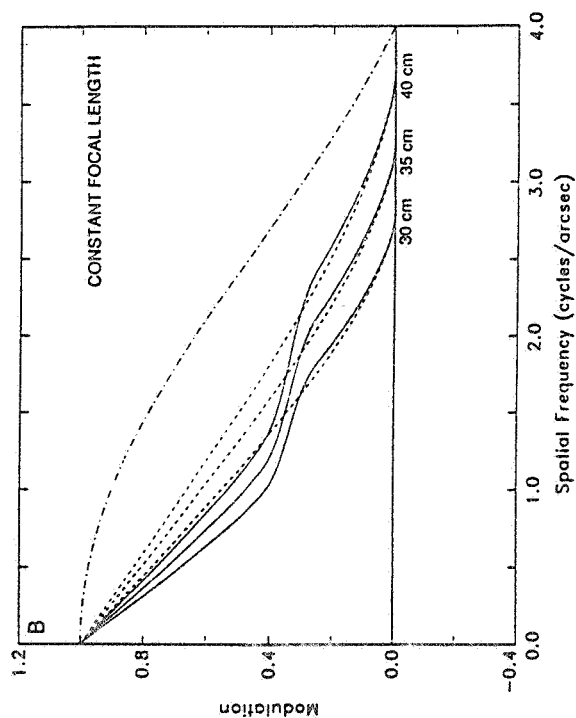
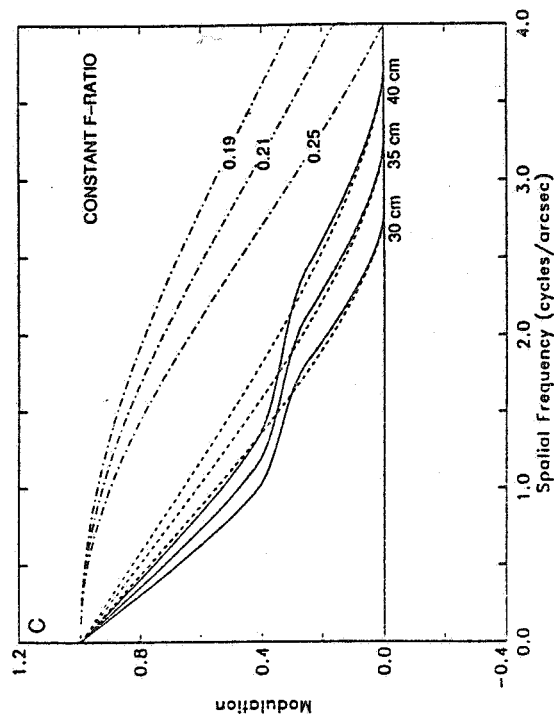
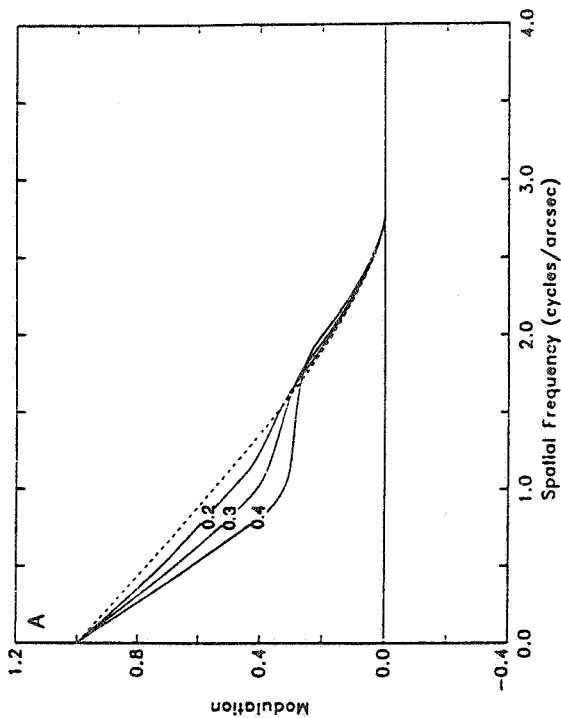


Figure 12. Image quality for vector magnetograph systems. The image quality for various optical systems is evaluated using the modulation transfer function (MTF) for each system. (a) MTF for a 30-cm telescope. The MTF's are shown for a 30-cm circular aperture (dashed line) and for 30-cm annular apertures with obscuration ratios of 0.2, 0.3, and 0.4 (solid lines). The annular apertures are equivalent to an aberration-free Cassegrain telescope. (b) MTF's for different apertures with a fixed focal length. The MTF's are shown for 30-, 35-, and 40-cm apertures with the same obscuration ratio of 0.3 (solid lines); the dashed curves show the MTF's for circular apertures with these same diameters. The MTF's for a detector array with 0.25 arcsec pixels is shown as a dashed-dotted curve. For the cases shown, the detector's pixel size in arcsecs remains fixed because the focal length is kept constant. (c) MTF's for different apertures with a fixed f-ratio. The MTF's are shown for annular apertures of 30, 35, and 40 cm with an obscuration ratio of 0.3 and a fixed f-ratio of 74.25. This f-ratio is the value determined by a 0.25 arcsec pixel with a physical dimension of 27 microns.

dimension depends on the focal length. The total fields of view for a 1024 by 2048 array are then 4.26×8.53 , 3.66×7.31 , and 3.20×6.40 arc min², respectively. For a fixed f-ratio and fixed array size, as the telescope aperture increases the spatial resolution increases but the field of view decreases. This trade-off has determined the aperture size by optimizing the system to have maximum view of field with maximum spatial resolution.

The combined modulation transfer function for a system of telescope and detector is shown in Figure 13 where the system MTF's for three telescope apertures of 30, 35, and 40 cm with an obscuration ratio of 0.3 are given. The f-ratio was held constant which allows scaling of the optical components. The detector pixel size is 27 microns. The equivalent Rayleigh limit of a 9% modulation gives the values of 0.47, 0.40, and 0.35 arc sec for the spatial resolutions of the systems. Thus, an increase of 30% in the aperture from 30-cm decreases the Rayleigh resolution by only 33%. Dramatic changes in resolution are obtained only when the aperture changes are on the order of 100%. Since there is no major increase in resolution in going from a 30-cm to a 40-cm system, and cost and weight factors argue for the smaller apertures, we have selected a 30-cm aperture for the optical system. We note in passing that an increased aperture will decrease the exposure time required but as will be seen in the detector section, this gain would be counteracted by the complexity of the detector electronics.

1.4 Instrument Design Summary

A schematic of the instrumental design for the vector magnetograph is shown in Figure 14a. The basic elements are: (1) the telescope optics which include a 30-cm Cassegrain telescope, heat-rejecting prefilter, articulated secondary, and relay-collimator lenses; (2) a high-precision polarimeter to provide the polarization analysis of magnetically sensitive photospheric lines by transmitting selected states of polarized light through to the detector; (3) a tunable filter to select specific spectral lines and transmit portions of the line profiles through to the detector system; and (4) a detector system that includes an intensity modulator to maintain incident intensities below CCD saturation levels, an imaging solid-state array, and associated electronics to digitize the polarized intensities measured over the 4.3×8.6 arc min field of view.

An H-alpha telescope will be coaligned with the SAMEX magnetograph. The H-alpha telescope will provide (1) imaging of the tracers of the chromospheric magnetic field (e.g., fibrils, filaments, and arch filament systems), (2) observations of chromospheric activity (e.g., flares and filament activations), (3) cospatial identification of optical emission with magnetic features seen in magnetograms (e.g., field structures in the network filigree and locations of umbra and penumbra), (4)

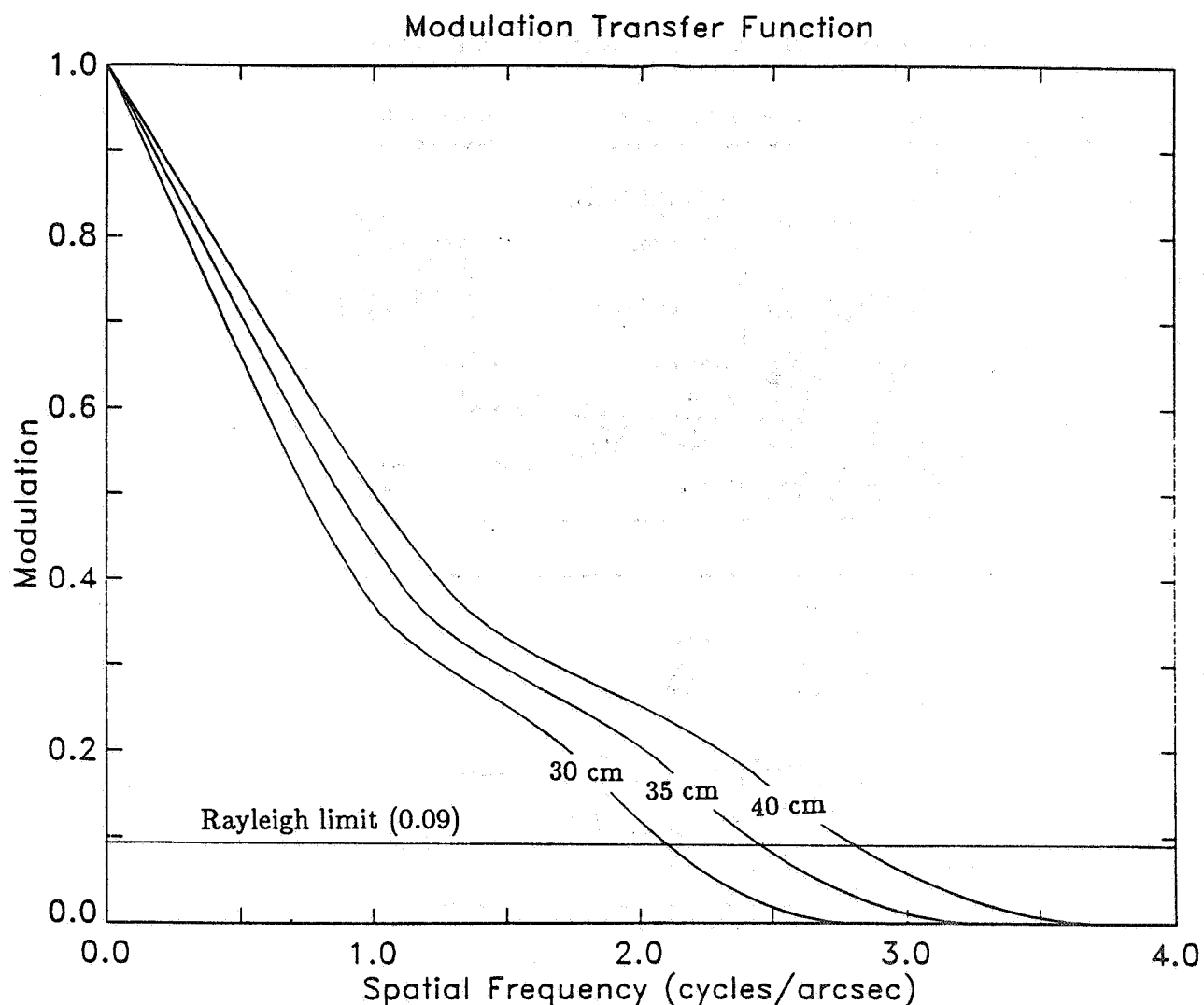


Figure 13. The resultant modulation transfer function for a telescope and detector system. The system MTFs for three telescope apertures of 30, 35, and 40 cm with obscuration ratio of 0.3 are shown. The f-ratio was held constant to allow scaling of the optical components. The size of the detector pixel was taken to be 27 microns. The equivalent Rayleigh limit of a 9% modulation gives the values of 0.47, 0.40, and 0.35 arc sec, respectively, for the spatial resolutions of the three systems. An increase of 30% in aperture size decreases the Rayleigh resolution by 33%. To significantly change the spatial resolution requires an increase of the aperture size on the order of 100%. These results show there is no dramatic increase in spatial resolution in going from a 30-cm aperture to a 40-cm system.

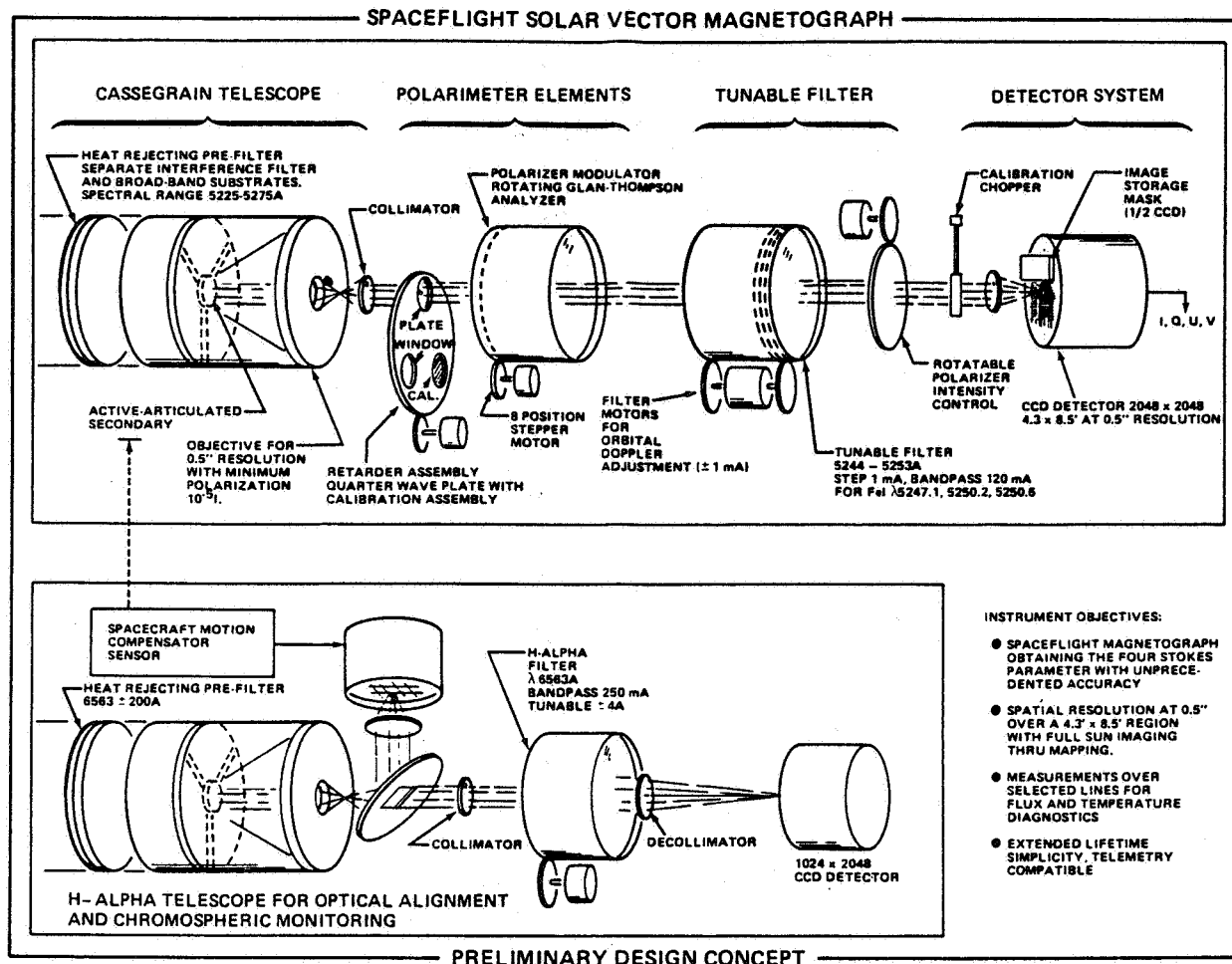


Figure 14. Schematic designs for the SAMEX magnetograph and H-alpha telescope. (a) The four basic elements of the magnetograph system. The telescope optics include a 30 cm Cassegrain telescope, heat-rejecting prefilter, articulated secondary mirror, and relay-collimating lenses. A high-precision polarimeter provides the polarization analysis of magnetically sensitive spectral lines by transmitting selected states of polarized light through to the filter and detector. A tunable filter selects the specific spectral line to be analyzed and transmits portions of the line profile through to the detector. The detector is envisaged as a 2048 x 2048 CCD array with half of the array used for imaging and half for temporary storage of an image while the image is being read out. (b) A schematic of the optical design for the H-alpha telescope. The basic elements are a 30-cm Cassegrain telescope, tunable H-alpha filter, 2048 x 2048 array detector, and the optics and sensor for the image motion compensator. The H-alpha telescope is to have the same resolution as that of the magnetograph and thus has a similar optical system.

chromospheric velocities , and (5) full-Sun imaging for compensation of spacecraft motion.

A schematic of the optical design for the H-alpha telescope is also shown in Figure 14b. The basic elements are a 30-cm Cassegrain telescope, tunable filter, detector, and sensor for the motion compensator. The H-alpha telescope will have the same resolution as that of the magnetograph. There are three arguments for having a separate 30-cm aperture for the H-alpha system instead of using one telescope. First, it is essential to avoid introducing any unnecessary polarization into the magnetograph optics through the introduction of a beam splitter. Second the image of the full Sun is needed for the image-motion compensator. Finally, the operating wavelength range of the individual telescopes can be limited and this will lead to better optical designs for both telescopes and better performances from coatings that are optimized for the narrow spectral range.

In the following sections, II.2 to II.6, each of the major components of this design for the SAMEX vector magnetograph will be described and discussed in detail. In the final section, II.7, a summary of the overall performance of the SAMEX magnetograph that is anticipated from this particular design is given.

2.0 The Optical System

2.1 Introduction

Our specific objectives were to design a magnetograph optical system with a wide field of view and a polarization sensitivity of 10^{-5} or less which is better than that of any existing magnetograph. These objectives necessitated a different approach to optical design methods. A Cassegrain telescope with aluminum coatings normally exhibits a polarization sensitivity on the order of 10^{-3} . The design of the SAMEX magnetograph optics employed the innovative "polarization aberration theory" (Chipman, 1987) which provided an increase of 4 orders of magnitude in optical polarization sensitivity compared with other systems. Also, compared with the only operational filter magnetograph (MSFC), the design has achieved 128 more resolvable elements and 8 times the angular resolution in its optical performance to provide coverage of an entire solar active region. Thus, in all aspects the system meets the SAMEX scientific objectives and will be dramatically superior to existing ground-based and spaceborne magnetographs.

The optical study also included: (1) the first demonstration that polarization aberration theory is valid, simple and effective in lens design, and (2) development of a whole new class of ultralow polarization lenses. Thus the objective of a superior magnetograph has been met and a major new paradigm for optical system design has been established.

2.2 Optics Overview

The design goal for the optics of the magnetograph is to determine the Stokes vector associated with light from an extended region of the Sun. This is done by producing a series of solar images with a narrow spectral bandwidth on an array detector in different polarization states without the introduction of noticeable instrumental polarization effects. From these Stokes measurements at several wavelengths, the solar magnetic field can be determined.

The functions of the optical elements of the magnetograph are:

1. to collect sunlight,
 2. to transmit incident light only in a specific polarization state,
 3. to filter the sunlight into a tunable 0.12 Å bandwidth,
- and
4. to image this sunlight onto the array detector.
- Of these requirements, polarizing and filtering the light provide the greatest technical challenges.

Table 9 lists the optical specifications associated with the magnetograph as shown in Figure 14a. An overview of the particular subassemblies of the magnetograph is presented in the following paragraphs.

Prefilter

The prefilter is a large blocking filter placed in front of the Cassegrain telescope. The prefilter limits the spectral range transmitted into the telescope to a 20-nm bandpass, reflecting the balance back into space. The prefilter keeps solar heat out of the telescope. It also limits the spectral range making the subsequent filtering easier. Since the prefilter is before the rest of the system, its range of angles of incidence is the same as the field of view of the telescope, a minuscule 5 minutes of arc, making its polarization effects negligible. However, the introduction of birefringence by thermal and mechanical stress must be avoided in the mounting and operation of the prefilter.

Table 9. Summary of optics performance for SAMEX solar magnetograph.

Spectral bandwidth	0.120 Å
Design wavelengths	5244-5253 Å
Field of view	4.3 x 8.5 arc minutes
Telescope aperture	30 cm
Angular resolution	0.47 arc sec
Wave front (over entire image)	>0.08 waves RMS
Instrumental polarization (optics and coatings)	<10 ⁻⁶
Focal length	23.8 meters
Focal length of primary	1.2 meters
Focal length of secondary	0.5495 meters
Overall length	2.3 meters
Detector array size	55.3 x 27.7 mm
Number of pixels	2048 x 1024 (imaging section)
Pixel size	27 x 27 micrometers
f-number	79.21
Airy disk diameter	44 micrometers
Total transmission	6.7%
Central obscuration	30%
Optical elements (before polarimeter)	2 mirrors, 6 lenses
Optical elements (after filter)	5 lenses

Cassegrain Telescope

The collecting optical system chosen is a Cassegrain telescope. The magnetograph has a field of view of 4.3 by 8.5 arc min. The Cassegrain design form is a size and cost efficient method to achieve diffraction limited image quality over the field of view. A Cassegrain telescope optical system was chosen because of the following properties inherent in a Cassegrain design:

1. a radially symmetric optical design which minimizes residual optical polarization;
2. no folding mirrors which would introduce substantial polarization effects;
3. small reflecting angles on the order of the large focal ratio (focal length/aperture diameter) which are minimized for an obscuration ratio (ratio of diameter of the secondary to the primary) of 1/3;
4. a nearly aplanatic (coma free) system due in part to the large effective focal length;
5. compact system with the total length about half the focal length of the primary whereas the effective focal length is given by the product of the magnification power of the secondary and the focal length of the primary;
6. less weight than a refracting system;
7. accessible focal plane behind the primary mirror;

8. focal plane fixed by focusing with the secondary (small tilts at the secondary provide large corrections at the image plane).

The Cassegrain selected is a Ritchey-Chretien design with hyperbolic primary and secondaries. This form provides the largest field of view of the various Cassegrain forms. Since the instrument is to be spaceborne, the optics will be diffraction limited to minimize size and weight.

A substantial problem associated with using a reflecting telescope is the polarization effects associated with the metallic reflections. The instrumental polarization from an ordinary Cassegrain telescope is about 30 times greater than the ultimate sensitivity of the design goal of the SAMEX polarimeter. These polarization effects are also much larger than those associated with an equivalent refractor. Large lenses are less practical for this system because of weight, fabrication, and cost considerations. So, to meet the requirements on instrumental polarization, a detailed polarization design study has been done on the Cassegrain. This study encompasses an optical design, thin film design, and polarization analysis. This is described in detail in the section on polarization analysis, section II.3. The conclusion is that the use of specially designed coatings on the telescope mirrors can reduce the instrumental polarization to levels far below the design goal in sensitivity of the polarimeter. Thus, the instrumental polarization of the Cassegrain telescope need not be the limiting factor in the accuracy of spaceborne solar magnetographs.

Not detailed in this study is the impact of light scattered from baffles and optical surfaces on the limiting polarization accuracy of the system. However, the polarization effects of scattered light can be controlled by proper baffling design and correct surface coatings. A complete baffle design and analysis of the system should be modeled as part of the next study phase using a stray light analysis program (e.g., Analysis Program Arizona Radiation Trace code, APART).

Active Articulated Secondary Mirror

To achieve a resolution of 0.5 arc sec will require holding the pointing accuracy to <0.1 arc sec with respect to the viewing target. To reduce image motion down to this level of 0.1 arc sec, any detected image motion will be compensated by moving the active secondary mirror of the telescope. The compensating movements of the secondary will be governed by signals from the image-motion compensator of the coaligned H-alpha telescope.

Star sensors on the host spacecraft will provide the roll compensation (~ 1.8 arc min/3 min). The solar limb sensor, located within the H-alpha instrument, and the star tracker system, which gives the third dimension of stability (roll), will

(1) provide the necessary corrections to the alignment actuators on the secondary for fine pointing to within 0.1 arc sec within the pointing ability of the host spacecraft (~10 arc sec) and (2) remove spacecraft drift and pointing errors. A feedback system of the tracking and vehicle control loops can be provided in conjunction with the control moment gyros. (The image motion compensator and secondary system alone might also allow mapping of the entire Sun by moving the secondary, but this technique would require additional study to determine how it might affect the optical performance.) Further consideration of the active articulated secondary mirror is given later in section III.2.3. Special attention to the optical design is required to avoid degradation of the optical system by the secondary offsets. The associated spacecraft support system is assumed to be able to point within 10 arc sec and hence the optical degradation produced by the articulation was not considered.

Relay Optics

The function of the relay optics is to produce an intermediate solar image of the correct size for a field stop, and then to prepare the optimum beam parameters for the polarimeter and tunable filter. In addition, the relay optics are designed to shorten the overall length of the system as much as possible and to have small angles of incidence to minimize instrumental polarization from lens coatings. Special lens coatings with very low polarization effects have been designed for the relay optics. These coatings are included in the calculations for the system's instrumental polarization. The intermediate focal plane within the relay section allows the placement of a field stop and shutter. A calibration of the distortion and registration of images can be accomplished by placing a fiducial screen at this focal plane (but this calibration does not include distortion effects of the primary).

Polarimeter

The function of the polarimeter is to transmit one polarization state and discard the orthogonal polarization state, and maintain the wavefront quality of the transmitted state so that the imaging quality of the instrument is not degraded. The polarimeter is discussed at length in section II.4, and only its optical integration into the system is covered here. The polarimeter consists of a turret-mounted retarder assembly followed by a rotating linear polarizer. The polarimeter is used in collimated light to eliminate spherical aberration and other aberrations which occur when spherical waves propagate through plane parallel blocks of optical material.

The retardance of a retarder varies with angle of incidence as a quadratic function. Thus it is desirable to minimize angles through the retarder and use special field-widened retarder designs which compensate for much of this effect. Light is collimated before the polarimeter so that the

retardance does not vary across the pupil for any given field point. This eliminates wave front aberrations from the retarder but does leave some small retardance variations which persist across the field of view. These variations will be calibrated and can be subtracted out during data reduction.

Glan-Thompson prisms are specified as the linear polarizers since these prisms have the best extinction ratios (10^{-6}) currently available. Glan-Thompson polarizers are comprised of two calcite prisms cemented together. One polarization state is transmitted straight through, whereas the orthogonal state undergoes total internal reflection and is absorbed on the side of the prism. The length-to-clear-aperture ratio of Glan-Thompson prisms is 2.5, making them long devices. Thus it is necessary to use them in a nearly collimated beam to avoid additional aberrations. An additional sheet polarizer is used to arrive at the 10^{-4} measured polarization accuracy (see the polarimeter section) without cross talk between the different states of polarization.

It is desirable to use a large beam and small angular field of view through the polarimeter to reduce the range of angles of incidence at the retarders. The diameter is limited to about 35 mm by the size of available Glan-Thompson polarizers.

A rotating quarter-wave plate is placed behind the rotating linear polarizer so that the transmitted linearly polarized light will not be modulated by any optical elements following the polarimeter that are polarizing elements, e.g., a birefringent filter.

Tunable Filter

In section II.5 the detailed designs of different types of filters are compared and discussed. In our optical design, the tunable filter is placed in collimated light with the aperture stop of the system, the primary mirror, imaged into the center of the filter. In section II.5, we discuss how this location distributes the effect of defects relatively evenly over the entire image. Locating the pupil in the center of the filter and balancing the pupil diameter against the angle of the chief ray will minimize the active volume of the tunable filter. Since a birefringent tunable filter is a long narrow device limited to about 30 mm clear aperture, it is best to transmit light through it with a relatively large pupil located at the center of the filter. This large pupil leads to a relatively small range of angles through the filter. This has the added benefit of reducing the wavelength shift with field which is a quadratic function of the angle of the chief ray through the filter.

Imaging Optics

The imaging optics image the collimated light emerging from the tunable filter onto the detector. Given the field of view of the system, 4.3 by 8.5 arc min, and the size of the detector, 55 x 55 mm², the effective focal length of the entire system needs to be 21 meters. The lens group following the filter requires a focal length of 1.3 meters to achieve this. To shorten the imaging optics from 1.3 meters but maintain a 1.3-meter effective focal length, a telephoto system was designed. The first lens group has a positive focal length and the second group a negative focal length. The entrance pupil for the imaging optics is in the center of the tunable filter, 200 mm in front of the lens. This complicates the aberration balancing and results in most of the lenses bending toward the entrance pupil.

Instrumental polarization is not an important consideration in the imaging optics. The light has already passed through the polarimeter, so polarization effects in the imaging optics do not greatly affect the polarimetric accuracy of the system. The effect of ghost images due to the optical system could be a minor problem and should be analyzed in the next study phase.

Color Correction

The entire optical system is color corrected over the spectral range 480 to 560 nm, the focal length being nearly constant over this range. This makes the system far more useful than a monochromatic design and will prove helpful during the assembly and testing of the instrument. The coatings are wavelength sensitive and do not perform well at the extreme limits of this extended 80 nm spectral range. The system is used only in a ~10 nm range.

The color correction is accomplished through the use of color corrected doublets and lens assemblies throughout the design. The cost and complexity of performing the color correction are small relative to its utility.

Folding Considerations

The system can be shortened by a set of folding mirrors placed after the spectral filter. This would reduce the system by approximately 50 cm, but due to the small effect (about 20% of the overall length), this feature was not incorporated in the current design. The folding of the system cannot be performed before the polarimeter since any folding mirrors would introduce severe polarization effects.

THE SYSTEM CONFIGURATION FOR THE SOLAR VECTOR MAGNETOGRAPH

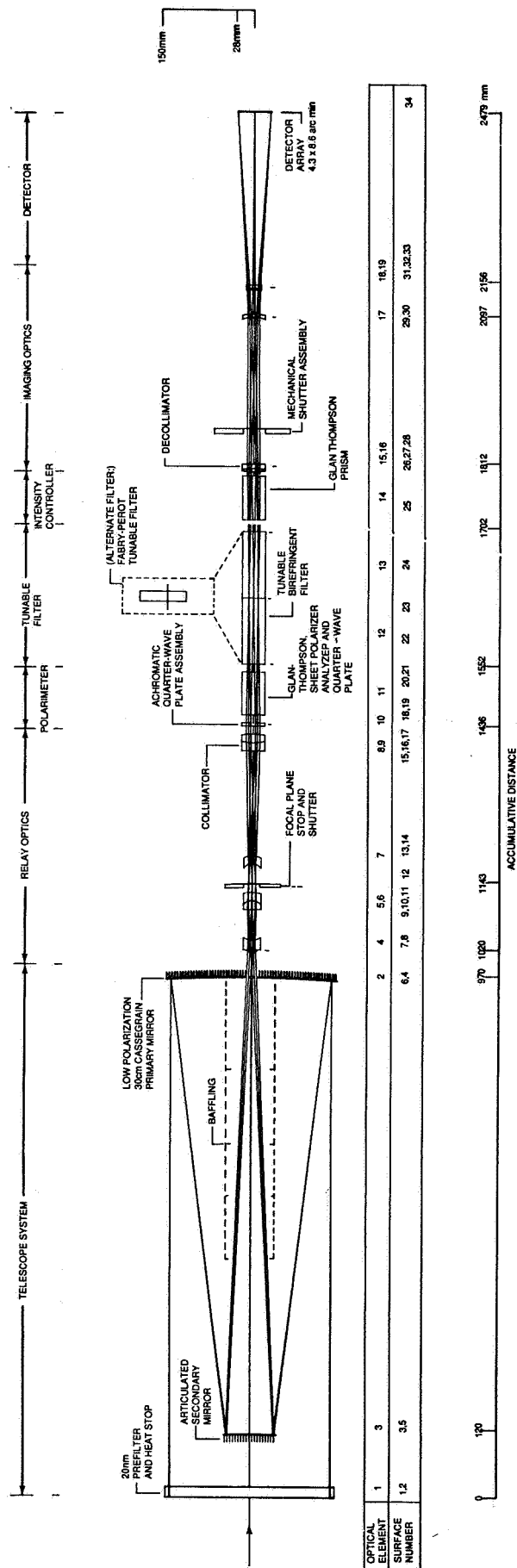


Figure 15. Detailed design of the optical system and subsystems for the SAMEX solar vector magnetograph. The major elements are the telescope, relay optics, collimator, polarimeter, spectral filter, imaging optics, and large-array detector. Each optical element and surface is numbered, and these numbers are located beneath the respective elements and surfaces on the figure. The elements E(1) are defined in Table 10 and the surfaces S(1) are defined in Table 13. The chief ray and extreme marginal rays have been drawn.

2.3 Optical Performance

This section describes the optical system designed for the SAMEX vector magnetograph and its optical performance. Figure 15 shows a scale drawing of the optical system with the various optical subsystems labeled. The analysis of the optical system proceeds as follows: (1) The imaging properties of the entire optical train are treated. (2) The imaging performances of the optics before and after the polarimeter and filter are analyzed separately. These optics have been designed to be diffraction limited separately to facilitate testing and alignment of the system. Follow-on studies should include: (1) an analysis of the aiming and tracking method of the system as it affects the optical performance, (2) design of a baffle system for control of stray light in the optics and an analysis of its effect on the polarization sensitivity, and (3) a thermal analysis of the solar heat load.

Optical Prescription of the SAMEX Optical System

The SAMEX optical system was designed using the CODE V optical design program running on a MicroVAX computer at the Lens Design Laboratory of The University of Alabama in Huntsville's Center for Applied Optics. Tables 10, 11, 12, and 14 describe the detailed results of this optical design. Table 10, the table of optical specifications, contains the curvatures, thicknesses, materials, and clear apertures of the optical elements. Table 11, the table of aspheric specifications, contains the optical prescriptions for the two non-spherical elements, the hyperbolic primary and secondary mirrors. Table 12 contains the first-order optical parameters of the system. Table 13, the table of CODE V specifications, contains the input file necessary to recreate the system model using CODE V. Table 14 contains the paraxial ray trace and Table 15 contains the calculation of the third-order transverse aberration coefficients surface by surface. The system aberrations are small, relative to a pixel, as shown by the transverse ray aberration sums given at the bottom of Table 15.

Table 10. Optical specifications for the SAMEX
spaceborne solar vector magnetograph.

ELEMENT NUMBER E(i)	RADIUS OF CURVATURE		THICKNESS	APERTURE RADIUS		GLASS
	FRONT	BACK		FRONT	BACK	
OBJECT	Infinity		Infinity			
Prefilter						
1	Infinity	Infinity	20.000000 950.000000	151.252365	151.235260	BK7
Primary Mirror, Aperture Stop						
2	Aspheric (1)		-850.000000	150.000000		Reflector
Secondary Mirror						
3	Aspheric (2)		900.000000	44.944843		Reflector
Relay Lenses						
4	115.76704	29.36186	19.084177 55.860219	6.982490	6.038970	BK7
5	96.25768	-22.59102	16.773049	8.016703	7.932736	KF9
6	-22.59102	-80.58434	15.000000 16.527328	7.932736	7.689602	FK5
Intermediate Solar Image			35.941797	6.508429		
Field Lens						
7	-19.64090	-30.56832	15.000000 195.611279	6.145518	7.658357	FK5
Collimating Lenses						
8	6347.81597	68.01464	15.000000	12.465919	12.714062	KF9
9	68.01464	-119.00923	16.915258 14.784900	12.714062	13.021261	FK5
Retarder Filter Wheel						
10	Infinity	Infinity	6.000000 15.000000	12.657876	12.565335	CAL
Rotating Linear Polarizer						
11	Infinity	Infinity	80.000000 15.000000	12.213658	10.979787	CAL

Table 10. Optical specifications for the SAMEX
spaceborne solar vector magnetograph (cont'd).

ELEMENT NUMBER E(i)	RADIUS OF CURVATURE		THICKNESS	APERTURE RADIUS		GLASS
	FRONT	BACK		FRONT	BACK	
Tunable Filter						
12	Infinity	Infinity	125.000000	10.628110	8.700187	CAL
13	Infinity	Infinity	125.000000	8.700187	10.011110	CAL
			15.000000			
Intensity Control Polarizer						
14	Infinity	Infinity	80.000000	10.362690	11.596217	CAL
			15.000000			
Telephoto Imaging Lenses						
15	131.32668	-109.66039	6.382536	11.960589	11.886025	FK5
16	-109.66039	525.48780	7.548267	11.886025	11.791454	KF9
			271.054056			
17	-53.16385	-82.67376	5.613246	10.303558	10.684854	KF9
			44.547059			
18	-283.47838	-29.83727	3.367947	12.019952	12.056709	KF9
19	-29.83727	277.95944	5.613246	12.056709	12.259303	FK5
Image Distance			321.926278			
IMAGE	Infinity		0.000000			

Notes:

Positive radius indicates the center of curvature is to the right.
Negative radius indicates the center of curvature is to the left.
Dimensions are given in millimeters.
Thickness is axial distance to next surface.

Reference Wavelength = 525.0 nm
Color Corrected for Spectral Region = 480.0-560.0 nm

Table 11. Aspheric specifications.

Aspheric Equation:
$$Z = \frac{c\rho^2}{1 + (1 - c^2(k + 1)\rho^2)^{1/2}}$$

ASPHERIC	Curvature c	Conic Constant k	Element
A(1)	-0.00041667	-1.108871	Primary Mirror
A(2)	-0.00090988	-6.014593	Secondary Mirror

Table 12. First-order parameters.

Object distance	Infinity
Effective focal length	-23763.0671 mm
Back focal length	321.9259 mm
F/no.	-79.2102
Overall length	2256.6244 mm
Image height	30.8985 mm
Semi-field angle	0.0745 deg (4.47 arc min)
Entrance pupil diameter	300.0000 mm
Entrance pupil distance	963.1590 mm
Exit pupil diameter	6.0889 mm
Exit pupil distance	-160.3752 mm

Table 13. CODE V Specifications.

Surface S(i)	Radius	Thickness	Glass (element number)	
OBJ:	INFINITY	INFINITY		
1:	INFINITY	20.000000	BK7_SCHOTT	(1)
2:	INFINITY	100.000000		
3:	INFINITY	850.000000		
4 STO:	-2400.00000	-850.000000	REFL	(2)
CON:				
K:	-1.108871	KC: 100		
5:	-1099.04487	850.000000	REFL	(3)
CON:				
K:	-6.014593	KC: 100		
6:	1.0000E+11	50.000000		
7:	115.76704	19.084177	BK7_SCHOTT	(4)
8:	29.36186	55.860219		
9:	96.25768	16.773049	KF9_SCHOTT	(5)
10:	-22.59102	15.000000	FK5_SCHOTT	(6)
11:	-80.58434	16.527328		
12:	INFINITY	35.941797		
13:	-19.64090	15.000000	FK5_SCHOTT	(7)
14:	-30.56832	195.611279		
15:	6347.81597	15.000000	KF9_SCHOTT	(8)
16:	68.01464	16.915258	FK5_SCHOTT	(9)
17:	-119.00923	14.784900		
18:	INFINITY	6.000000	CALCITE	(10)
19:	INFINITY	15.000000		
20:	INFINITY	80.000000	CALCITE	(11)
21:	INFINITY	15.000000		
22:	INFINITY	125.000000	CALCITE	(12)
23:	INFINITY	125.000000	CALCITE	(13)
24:	INFINITY	15.000000		
25:	INFINITY	80.000000	CALCITE	(14)
26:	INFINITY	15.000000		
27:	131.32668	6.382536	FK5_SCHOTT	(15)
28:	-109.66039	7.548267	KF9_SCHOTT	(16)
29:	525.48780	271.054056		
30:	-53.16385	5.613246	KF9_SCHOTT	(17)
31:	-82.67376	44.547059		
32:	-283.47838	3.367947	KF9_SCHOTT	(18)
33:	-29.83727	5.613246	FK5_SCHOTT	(19)
34:	277.95944	321.926278		
IMG:	INFINITY	0.000000		

Table 13. CODE V Specifications (cont'd).

SPECIFICATION DATA

EPD	300.00000			
DIM	MM			
WL	560.00	525.00	480.00	
REF	2			
WTW	1	3	1	
INI	JDT			
XAN	0.00000	0.00000	0.00000	
YAN	0.00000	0.05220		0.07450
VUY	0.00000	0.00000	0.00000	
VLY	0.00000	0.00000	0.00000	

APERTURE DATA/EDGE DEFINITIONS

CA				
CIR S1		151.25	CIR S2	151.23
CIR S3	OBS SMO	46.60	CIR S3	151.10
CIR S4	OBS PM	10.00	CIR S4	150.00
CIR S5		44.94	CIR S6	PCA 10.00
CIR S7		6.98	CIR S8	6.03
CIR S9		8.01	CIR S10	7.93
CIR S11		7.68	CIR S12	6.50
CIR S13		6.14	CIR S14	7.65
CIR S15		12.46	CIR S16	12.71
CIR S17		13.02	CIR S18	12.65
CIR S19		12.56	CIR S20	12.21
CIR S21		10.97	CIR S22	10.62
CIR S23		8.70	CIR S24	10.01
CIR S25		10.36	CIR S26	11.59
CIR S27		11.96	CIR S28	11.88
CIR S29		11.79	CIR S30	10.30
CIR S31		10.68	CIR S32	12.01
CIR S33		12.05	CIR S34	12.25

REFRACTIVE INDICES

GLASS CODE	560.00	525.00	480.00
BK7_SCHOTT	1.518032	1.519868	1.522829
KF9_SCHOTT	1.524959	1.527289	1.531102
FK5_SCHOTT	1.488550	1.490126	1.492655

Note: No solves defined in system

Table 14. Paraxial ray trace for a wavelength of 525 nm. The marginal ray is parallel to the optical axis and strikes the edge of the primary (150 mm) mirror. The chief ray strikes the primary mirror at an angle of 0.0013 radians (4.46 arc min).

	Marginal Ray			Chief Ray		
	Height (mm) HMY	Angle (rad) UMY	[Index][Angle of Incidence] N * IMY	Height (mm) HCY	Angle (rad) UCY	[Index][Angle of Incidence] N * ICY
EP	150.000000	0.000000		0.000000	0.001300	
1	150.000000	0.000000	0.000000	-1.252368	0.001300	0.001300
2	150.000000	0.000000	0.000000	-1.235257	0.001300	0.001300
3	150.000000	0.000000	0.000000	-1.105230	0.001300	0.001300
STO	150.000000	0.125000	-0.062500	0.000000	-0.001300	0.001300
5	43.750000	-0.045385	-0.085193	1.105230	0.003312	0.002306
6	5.172397	-0.045385	-0.045385	3.920028	0.003312	0.003312
7	2.903126	-0.038439	-0.020308	4.085605	-0.009893	0.038603
8	2.169548	-0.020009	0.053881	3.896813	0.053960	0.186677
9	1.051826	-0.016874	-0.009082	6.911017	0.010543	0.125757
10	0.768802	-0.018143	-0.077747	7.087852	0.002981	-0.463080
11	0.496653	-0.030056	-0.036220	7.132568	-0.038939	-0.127450
12	-0.000101	-0.030056	-0.030056	6.489007	-0.038939	-0.038939
13	-1.080385	-0.038263	0.024950	5.089463	0.059099	-0.298065
14	-1.654331	-0.030492	0.023628	5.975948	-0.007752	-0.203247
15	-7.618830	-0.019550	-0.031692	4.459557	-0.005318	-0.007050
16	-7.912082	-0.022939	-0.207527	4.379784	-0.003845	0.090227
17	-8.300100	0.000001	0.069744	4.314746	-0.023499	-0.059755
18	-8.300084	0.000001	0.000001	3.967313	-0.015461	-0.023499
19	-8.300079	0.000001	0.000001	3.874545	-0.023499	-0.023499
20	-8.300063	0.000001	0.000001	3.522058	-0.015461	-0.023499
21	-8.300005	0.000001	0.000001	2.285152	-0.023499	-0.023499
21	-8.299988	0.000001	0.000001	1.932665	-0.015461	-0.023499
23	-8.299898	0.000001	0.000001	0.000000	-0.015461	-0.023499
24	-8.299807	0.000001	0.000001	-1.932665	-0.023499	-0.023499
25	-8.299791	0.000001	0.000001	-2.285153	-0.015461	-0.023499
26	-8.299733	0.000001	0.000001	-3.522058	-0.023499	-0.023499
27	-8.299716	0.020788	-0.063198	-3.874546	-0.006066	-0.053002
28	-8.167037	0.018470	0.141955	-3.913262	-0.006787	0.044137
29	-8.027622	0.020154	0.004877	-3.964489	-0.014343	-0.021888
30	-2.564891	-0.003461	0.068399	-7.852269	-0.060384	0.133356
31	-2.584316	0.011197	0.042456	-8.191217	-0.039980	0.059099
32	-2.085513	0.004792	0.018554	-9.972221	-0.038322	-0.004802
33	-2.069375	0.006641	0.113244	-10.101289	-0.030835	0.458528
34	-2.032099	0.006312	-0.000998	-10.274373	-0.064065	-0.101028
IMG	0.000000	0.006312		-30.898452	-0.064065	

Table 15. Third-order aberration coefficients for a wavelength of 525 nm.
The color coefficients were determined at 490 and 560 nm.

	SA ¹	TCO ²	TAS ³	SAG ⁴	PTB ⁵	DST ⁶	AX ⁷	LAT ⁸	PTZ ⁹
1	.000000	.000000	.000000	.000000	.000000	.000015	.000000	.097539	.000000
2	.000000	.000000	.000000	.000000	.000000	-.000015	.000000	-.097539	.000000
3	.000000	.000000	.000000	.000000	.000000	.000000	.000000	.000000	.000000
4	5.801535	-.362091	.005022	.000000	-.002511	.000000	.000000	.000000	.000833
	-6.433152	.000000	.000000	.000000	.000000	(aspheric contribution)*			
5	-2.002429	.162598	.001082	.004016	.005483	.000109	.000000	.000000	-.001820
	2.629564	.199287	.005034	.001678	.000000	(aspheric contribution)			
6	.000000	.000000	.000000	.000000	.000000	.000000	.000000	.000000	.000000
7	-.001906	.010868	-.011755	.002017	.008903	-.003834	-.029484	.056046	-.002955
8	-.002635	-.027389	-.129995	-.066733	-.035102	-.231206	-.058460	-.202540	.011649
9	-.000062	.002558	-.024614	-.001000	.010807	.013844	-.006087	.084288	-.003587
10	.000415	.007416	.046350	.016902	.002178	.100674	.012003	.071493	-.000723
11	.000923	.009741	.046577	.023725	.012299	.083484	.007850	.027624	-.004082
12	.000000	.000000	.000000	.000000	.000000	.000000	.000000	.000000	.000000
13	.000233	-.008360	.049413	-.017170	-.050461	.205113	-.011764	.140534	.016746
14	-.000352	.009088	-.045752	.006364	.032422	-.054745	.017058	-.146736	-.010760
15	.010723	.007156	.001756	.000694	.000164	.000154	.153859	.034224	-.000054
16	-.069999	.091301	-.040418	-.013955	-.000723	.006067	-.329728	.143357	.000240
17	.049234	-.126546	.116749	.044468	.008328	-.038099	.252629	-.216444	-.002764
18	.000000	.000000	-.000001	.000000	.000000	.004838	-.000005	.097541	.000000
19	.000000	.000000	.000001	.000000	.000000	-.004838	.000005	-.097541	.000000
20	.000000	.000000	-.000001	.000000	.000000	.004838	-.000005	.097541	.000000
21	.000000	.000000	.000001	.000000	.000000	-.004838	.000005	-.097540	.000000
22	.000000	.000000	-.000001	.000000	.000000	.004838	-.000005	.097540	.000000
23	.000000	.000000	.000000	.000000	.000000	.000000	.000000	.000000	.000000
24	.000000	.000000	.000001	.000000	.000000	-.004838	.000005	-.097538	.000000
25	.000000	.000000	-.000001	.000000	.000000	.004838	-.000005	.097538	.000000
26	.000000	.000000	.000001	.000000	.000000	-.004838	.000005	-.097537	.000000
27	.036627	.092154	.084834	.033309	.007547	.027936	.228906	.191977	-.002505
28	-.024211	-.022583	-.007470	-.002789	-.000449	-.000867	-.232812	-.072386	.000149
29	.000122	-.001641	.005387	.000476	-.001980	-.002135	.024948	-.111963	.000657
30	-.021309	-.124640	-.262578	-.100571	-.019568	-.196083	-.111791	-.217957	.006494

Table 15. Third-order aberration coefficients for a wavelength of 525 nm.
The color coefficients were determined at 490 and 560 nm (cont'd.).

	SA ¹	TCO ²	TAS ³	SAG ⁴	PTB ⁵	DST ⁶	AX ⁷	LAT ⁸	PTZ ⁹
31	.004968	.020745	.041460	.022209	.012583	.030914	.069916	.097322	-.004176
32	-.000458	.000356	-.003762	-.003700	-.003670	.000958	-.024657	.006382	.001218
33	.002773	.033685	.138041	.047113	.001649	.190763	.047059	.190544	-.000547
34	.000000	.000090	.005581	-.000517	-.003566	-.052279	-.000885	-.089594	.001183
SUM	-.019395	-.026207	.020941	-.003463	-.015665	.080592	.008562	-.013828	.005199

*(There are two components of aberrations at each surface: spherical and aspheric. Aspheric contributions occur only if the surface has a conic constant or fourth power deformation term.)

- ¹SA = Spherical aberration
- ²TCO = Coma
- ³TAS = Tangential astigmatism
- ⁴SAG = Sagittal astigmatism
- ⁵PTB = Petzval radius
- ⁶DST = Distortion
- ⁷AX = Axial color
- ⁸LAT = Lateral color
- ⁹PTZ = Petzval ratio

Optical Performance of the SAMEX Optical System

The optical design for the SAMEX vector magnetograph forms a diffraction limited image over the entire detector to keep the length of the overall system to a minimum. Figure 16 shows the diffraction MTF in cycles per arc seconds for objects located on axis, at 70% of the field of view, and at the edge of the detector (full field of view). The corresponding scale in terms of resolution in angular dimensions (arc min) and linear dimensions on the Sun (Mm) is given above the MTF. The diffraction limited MTF for an ideal optical system of the same aperture is also shown for reference. The cutoff frequency of the system is at 1 cycle per mm ($1/0.38 \text{ arc sec}^{-1}$). At 0.5 arc sec resolution, the modulation is greater than 0.15 across the entire image, in agreement with the theoretical MTF. The dip in the diffraction limited MTF at the middle frequencies is caused by the central obscuration of the Cassegrain telescope and is a result of the diffraction pattern of a circular aperture obscured by a secondary. The images are all diffraction limited with only a small drop in modulation evident in one direction at the full field of view; this is due to 0.08 waves RMS of astigmatism. The MTF for the imaging optics subassembly is effectively diffraction limited.

Figure 17 shows the point spread function of the optics for an object at the edge of the field of view (worst case). The point spread function is virtually an Airy disk ($1.22 \lambda/D$) because of the very small amount of aberration. Figure 18 is a plot of the wave front aberration showing the deviation of the transmitted wave front from a spherical wave front for the object at the edge of the field of view. The small amount of astigmatism is evident as a cylindrical wave front error.

Figure 19 contains the wave front aberration plots for fans of rays passing horizontally and vertically through the entrance pupil. Only one half the horizontal fan is shown since it is symmetric for a vertically oriented object. Wave fronts are plotted for an on-axis object, an object located at 70% of full field, and an object located at the edge of the field. Wave fronts are shown for 490 nm and 560 nm as well as 525 nm to demonstrate the color correction of the optics. The optical performance is close to diffraction limited across this entire spectral range due to the use of color-corrected lens pairs in the design. Figure 20 shows the transverse ray aberrations for the same fans of rays. The transverse ray aberrations are proportional to the derivative of the wave front aberrations. The majority of the aberration at 525 nm is within a 27-micron pixel. Figure 21 contains the monochromatic spot diagrams of the system with the Airy disk diameter and pixel size superposed for reference. This figure further demonstrates that the optical system is a diffraction limited system.

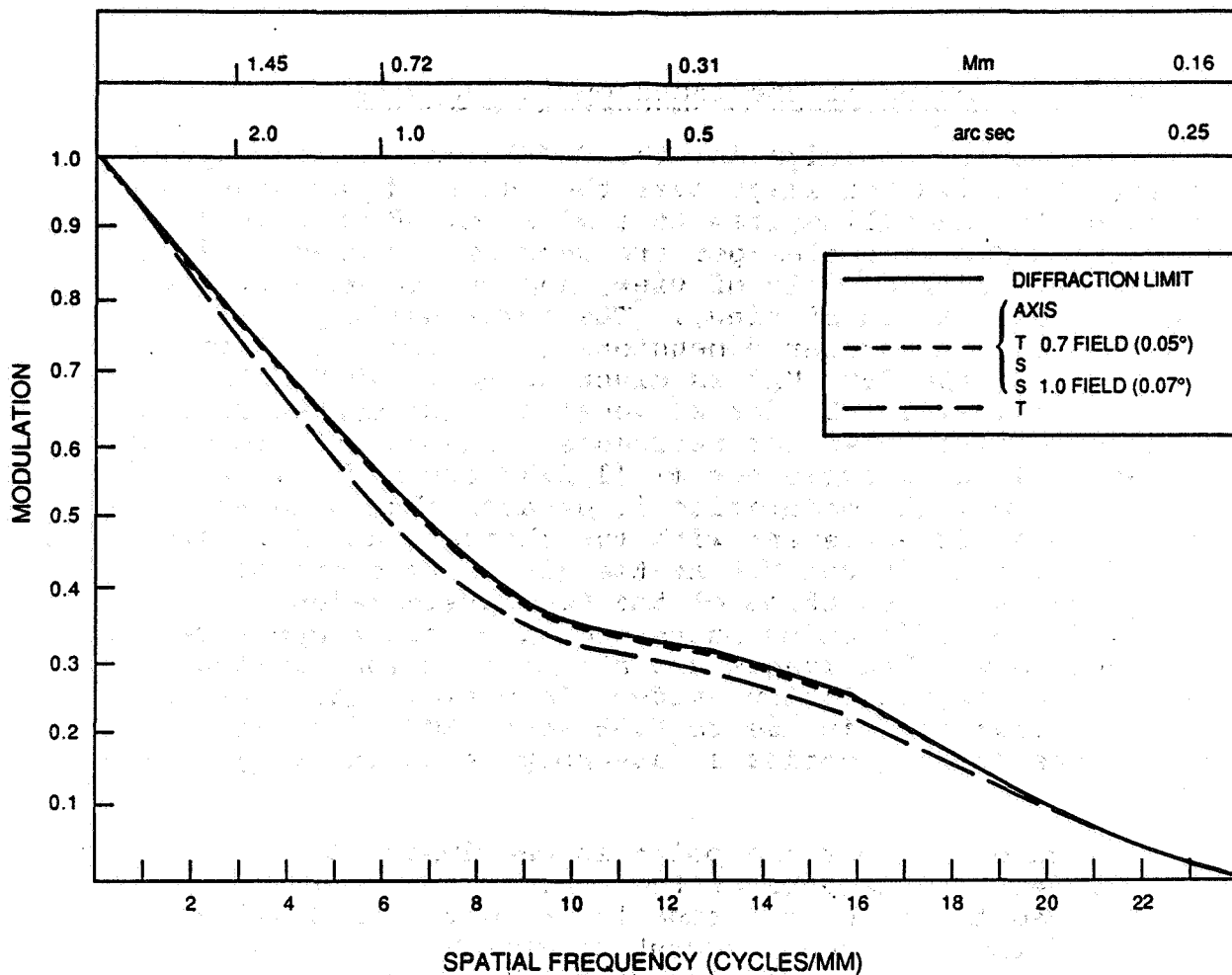


Figure 16. The diffraction MTF for the SAMEX optics. The diffraction MTF in cycles per arc sec is given for objects located on axis, at 70% from the center of the field of view, and at the edge of the detector (full field of view, equal to 256" or 0.07° from the center of the field). The corresponding scale in terms of resolution on the Sun in angular dimensions (arc sec) and linear dimensions (megameters, Mm) is given above the figure. The curves for the cases of on-axis, 70% full field (both tangential, T, and sagittal, S, directions), and full field in the sagittal direction all coincide and are indicated by the smaller dashes. The curve for the case of full field in the tangential direction is indicated by the longer dashes. The solid curve represents the diffraction-limited response.

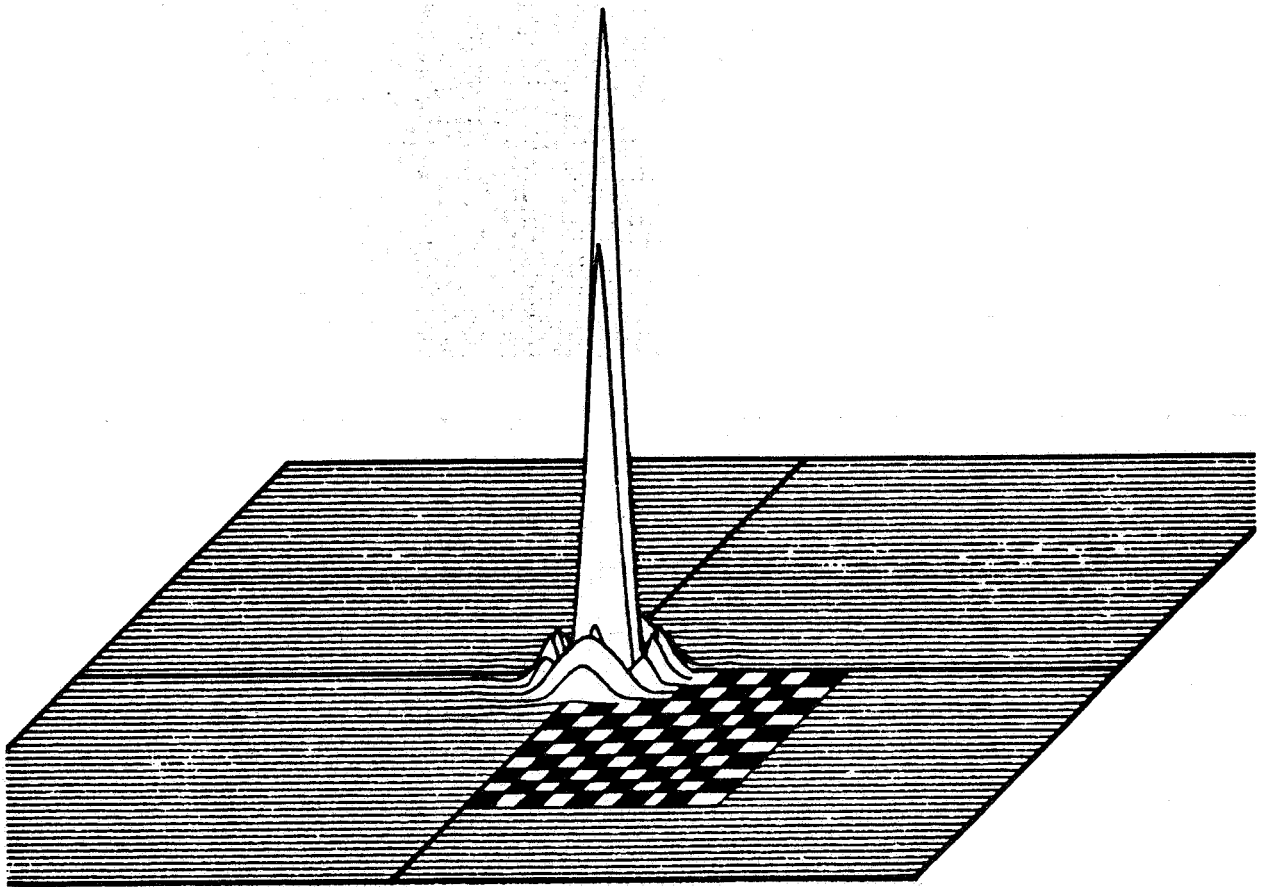
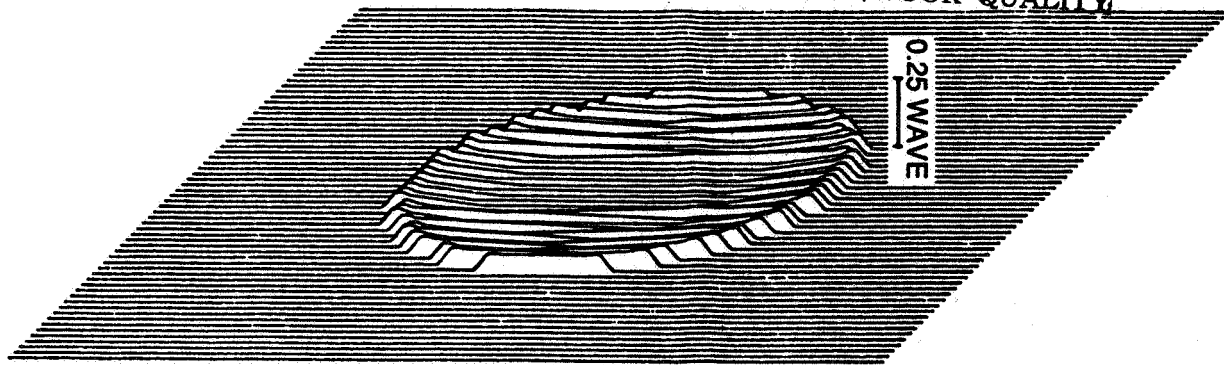
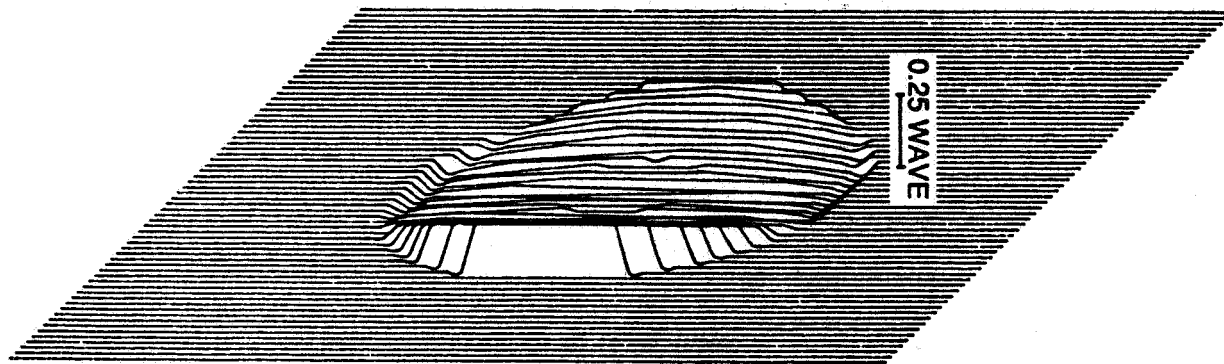


Figure 17. Point spread function of the SAMEX optics. This figure shows the point spread function of the optics for an object at the edge of the field of view which represents the worst case. The length scale is indicated by the checkerboard pattern: each square corresponds to a 27- μm pixel. The point spread function is virtually an Airy disk because of the very small amount of aberration.

ORIGINAL PAGE IS
OF POOR QUALITY



A.



B.

Figure 18. Wave front aberration of the SAMEX optics. This figure shows the deviation of the transmitted wave front from a spherical surface for an object located (a) on the optical axis and (b) at the edge of the field of view. The small amount of astigmatism is evident as a cylindrical wave front error. In both cases the wave front aberration is less than a quarter wave.

ORIGINAL PAGE IS
OF POOR QUALITY

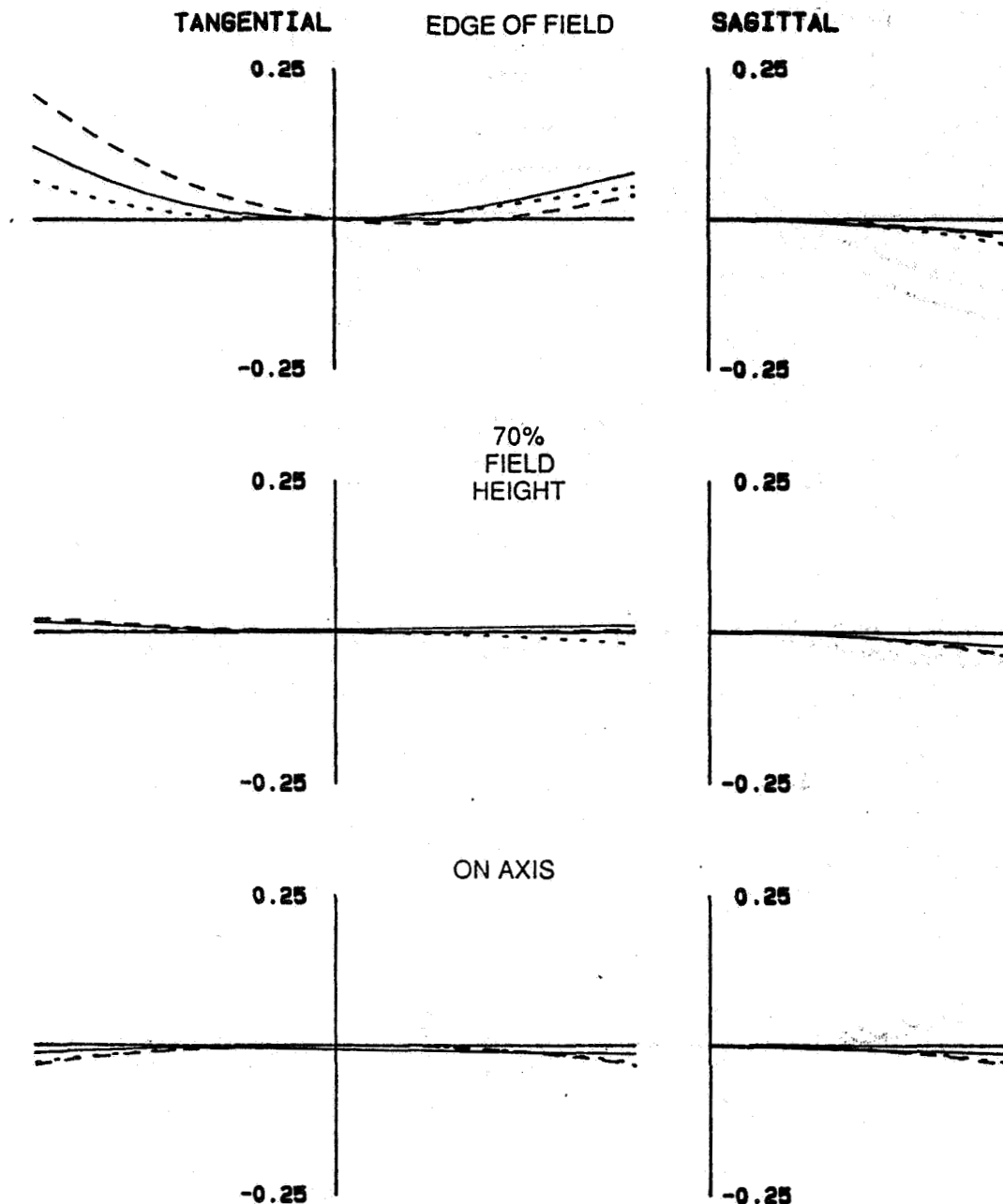


Figure 19. Polychromatic wave front aberrations of the SAMEX optics. These curves show the wave front aberrations of rays passing through the entrance pupil as a function of distance along the pupil's vertical (tangential) and horizontal (sagittal) axes for different wavelengths of light. Only half of the horizontal axis direction is shown since the wave front aberration is symmetric for a vertically oriented object. Wave front aberrations are plotted for an object at the far edge of the field of view, i.e., at 256" or 0.07° from the center of the field (top panels), at a distance of 70% of the full field of view (center panels), and on the optic axis (bottom panels). The curves in each panel show the aberrations for wavelengths of 4800 Å (dashed curve), 5250 Å (solid curve), and 5600 Å (dotted curve) to demonstrate the degree of color correction designed into the optics. The optical performance is close to a diffraction-limited system across this entire spectral range due to the use of color-corrected lens pairs in the design.

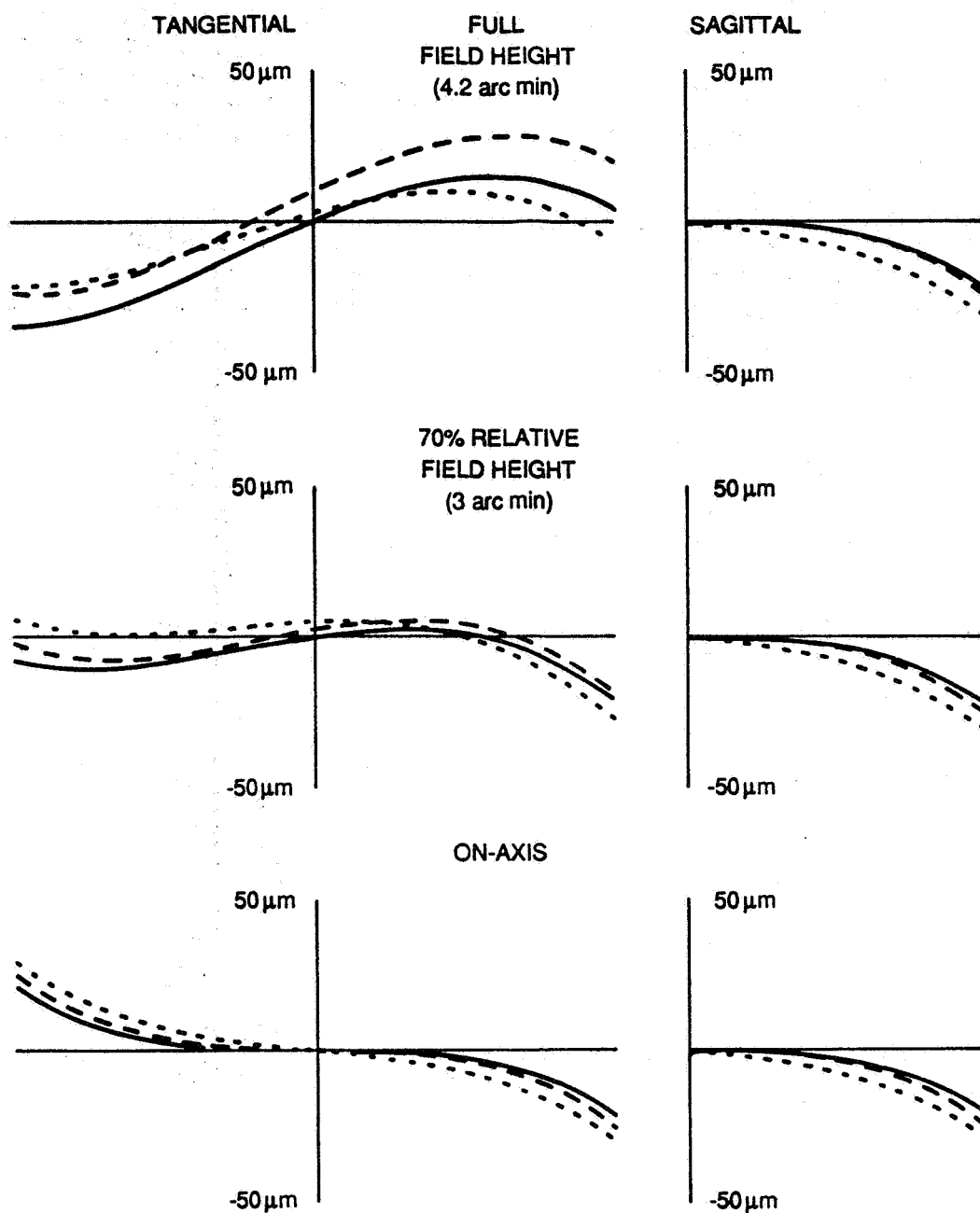
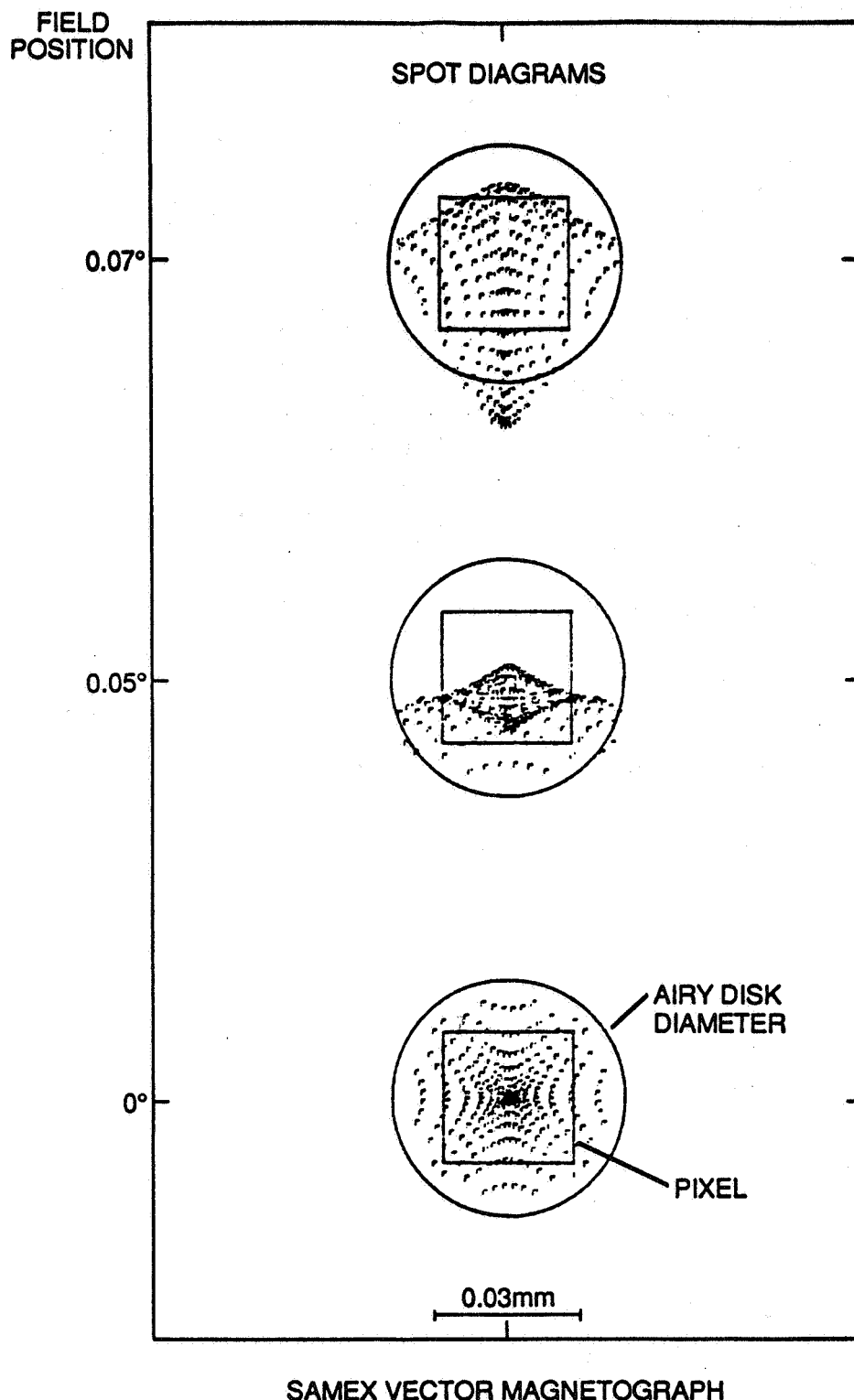


Figure 20. Transverse ray aberrations of the SAMEX optics. This figure shows the transverse ray aberrations for the same rays and object locations of Figure 19. The transverse ray aberrations are proportional to the derivative of the wave front aberrations.



SAMEX VECTOR MAGNETOGRAPH

Figure 21. Spot diagrams of the SAMEX optics. This figure shows the spot diagrams of the optical system for different positions in the field of view relative to the center of the field on the optical axis. The Airy disk and pixel size have been superposed for reference.

The distortion of the optics is shown in Figure 22 to be less than 0.5%. This corresponds to a nonlinearity in mapping of less than 5 pixels from the center to the corner of the detector. The astigmatic field curves are also shown in Figure 22. These are the longitudinal defocus for the sagittal and tangential fans as a function of field height. The difference between the two curves is the separation of the astigmatic foci. In this design the effect of the field curvature has been minimized by the standard method of flattening the sagittal field.

The optical performance for the SAMEX magnetograph has been designed to the diffraction limit with a resolution of 0.5 arc sec across the whole field of view, including the response function of the array detector. The distortion of the image at the detector is less than 0.5%. The system has been optimized for overall length, minimum angles of incidence, diffraction, spatial resolution, and field of view. The aberration analysis, spot diagrams, and modulation analysis all provide an evaluation of the optical design which shows the system reaches its design goal of 0.5 arc sec. The design is also compatible with the desire for minimum instrumental polarization as discussed in the next section.

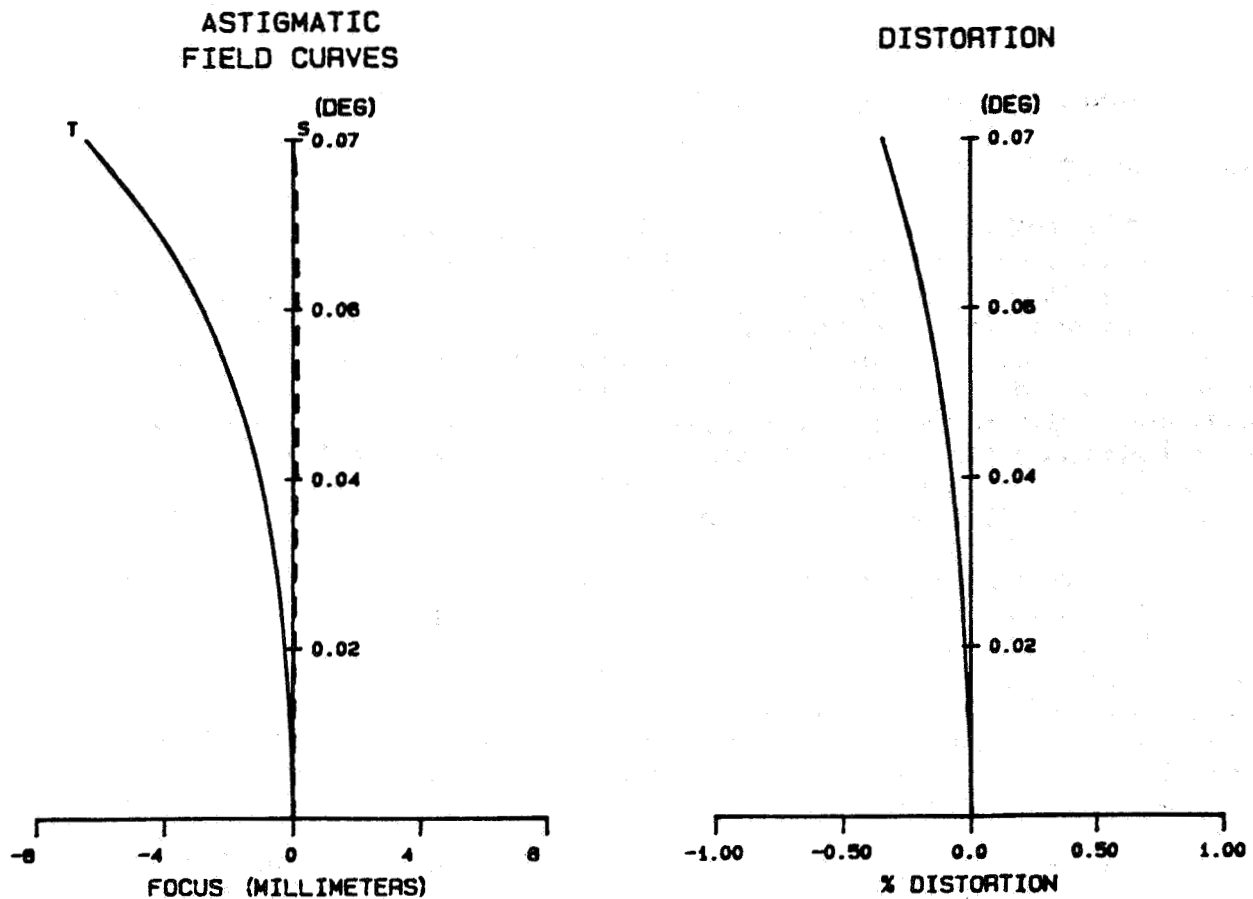


Figure 22. Distortion and astigmatic field curves for the SAMEX optics. The percent distortion and change in focus are shown for different locations in the field of view. The distortion data indicate that the distortion is less than 0.5% at the far edge of the field (256" or 0.07°). This corresponds to a nonlinearity in the mapping of less than 1.28" or about 5 pixels from the center to the farthest edge of the detector.

3. Analysis of Instrumental Polarization of Imaging Optics

3.1 Introduction

The objective of the SAMEX magnetograph's optical system is to accurately measure the polarization state of sunlight in a narrow spectral bandwidth over the field of view of an active region to make an accurate determination of the magnetic fields in that region. Our design goal is to measure magnetic fields to an accuracy of one part in 10^4 . To achieve this accuracy requires a polarimetric accuracy of 10^{-4} in determining the polarization components of the light as a fraction of the total intensity. This requirement means that the instrumental polarization of the optics must be reduced to levels below 10^{-5} .

All optical elements introduce some polarization change, especially when used off axis. Combinations of mirrors and antireflection-coated lenses can display a full range of polarization behavior: linear and circular polarization and linear and circular retardance. Of particular concern in the development of highly accurate polarimeters is polarization rotation which causes linearly polarized light to leak through subsequent crossed polarizers.

Because of this instrumental polarization, the SAMEX foreoptics (the optical elements in front of the polarizer - the Cassegrain telescope and the relay lenses) must be considered as a weak polarizer in front of the polarimeter section. This instrumental polarization in the foreoptics changes the polarization state of the sunlight incident on the polarimeter and thus introduces errors in the measurement of the solar magnetic field.

If possible, it would be preferable to place the polarimeter before the imaging optics, to locate polarizers and retarders in front of the Cassegrain telescope. Then the determination of the polarization state of light would be unaffected by the instrumental polarization of the foreoptics. This design would have the additional advantage of allowing the light to pass through the polarimeter over a smaller range of angles of incidence. However, such a design is impractical for a system with a 30-cm aperture - high quality polarizers and retarders have much smaller apertures, on the order of a few centimeters. Thus, it is necessary to use collecting optics to collect the 30-cm aperture of light and focus it through small polarizing elements in the polarimeter section. As the collected light is passed through smaller apertures, the angular spectrum of the light increases in a relationship governed by the Lagrange invariant. By reducing the beam from 30 to 2 cm through the tunable filter and polarimeter, the range of angles of incidence is increased by a factor of 15, from 5 to 75 min of arc, thereby increasing the angle of incidence effects in the lenses just in front of the polarimeter and in the polarimeter itself. Consequently we anticipate there will be some instrumental polarization due to the foreoptics in the light entering the polarimeter.

The goal of the analyses outlined in this section is to precisely characterize the extent of this instrumental polarization and to design the optics and coatings in unison to minimize this spurious polarization introduced by the foreoptics. We will show calculations of the instrumental polarization of ordinary foreoptics compared with the polarization performance of special ultra-low polarization optical coatings designed for this application.

The instrumental polarization analysis uses a program that incorporates the theory of polarization into the standard geometrical optics and lens design codes (Chipman, 1987). By including the polarization of the optical elements in the first- and third-order design process, the effects of coatings on curved substrates can be treated. For each ray, a polarization matrix (in the Jones matrix formulation) is calculated for the ray at each optical interface. These matrices are multiplied together to calculate the polarization matrix for that ray from object space to image space. An analytic function for the rays provides the polarization behavior as a function of the exit pupil, object height, and wavelength. This technique represents a quantum jump improvement in the practical design of foreoptics in front of a polarimeter and has led to a reduction in the degree of residual instrumental polarization to less than 10^{-5} in the SAMEX magnetograph system.

3.2 Instrumental Polarization of Standard Cassegrain Telescopes

To establish the need for a detailed polarization analysis, we estimate the polarization effects associated with a Cassegrain telescope with aluminum thin film coatings. The electric vector of light incident on a surface of the Cassegrain can be decomposed into two parts, the components of the vector vibrating parallel (p) and perpendicular (s) to the plane of incidence. These components have different reflections as a function of the angles of incidence. Figure 23a shows the reflectance for s and p polarized light from an aluminum thin film coating with complex index of refraction $n = 0.7 - 7.0 i$. Figure 23b shows the phase change on reflection for the s and p components. The difference in s and p reflectance causes a weak linear polarization aligned perpendicular to the plane of incidence to be associated with reflection from all metallic interfaces. Moreover, the differences in phase change cause a weak linear retardance to be associated with reflection from mirrors.

The percent reflections from the aluminum surface for the surface parallel (s) and surface perpendicular (p) rays are approximately given by

$$R_s = 0.94 + 0.02i^2,$$

and

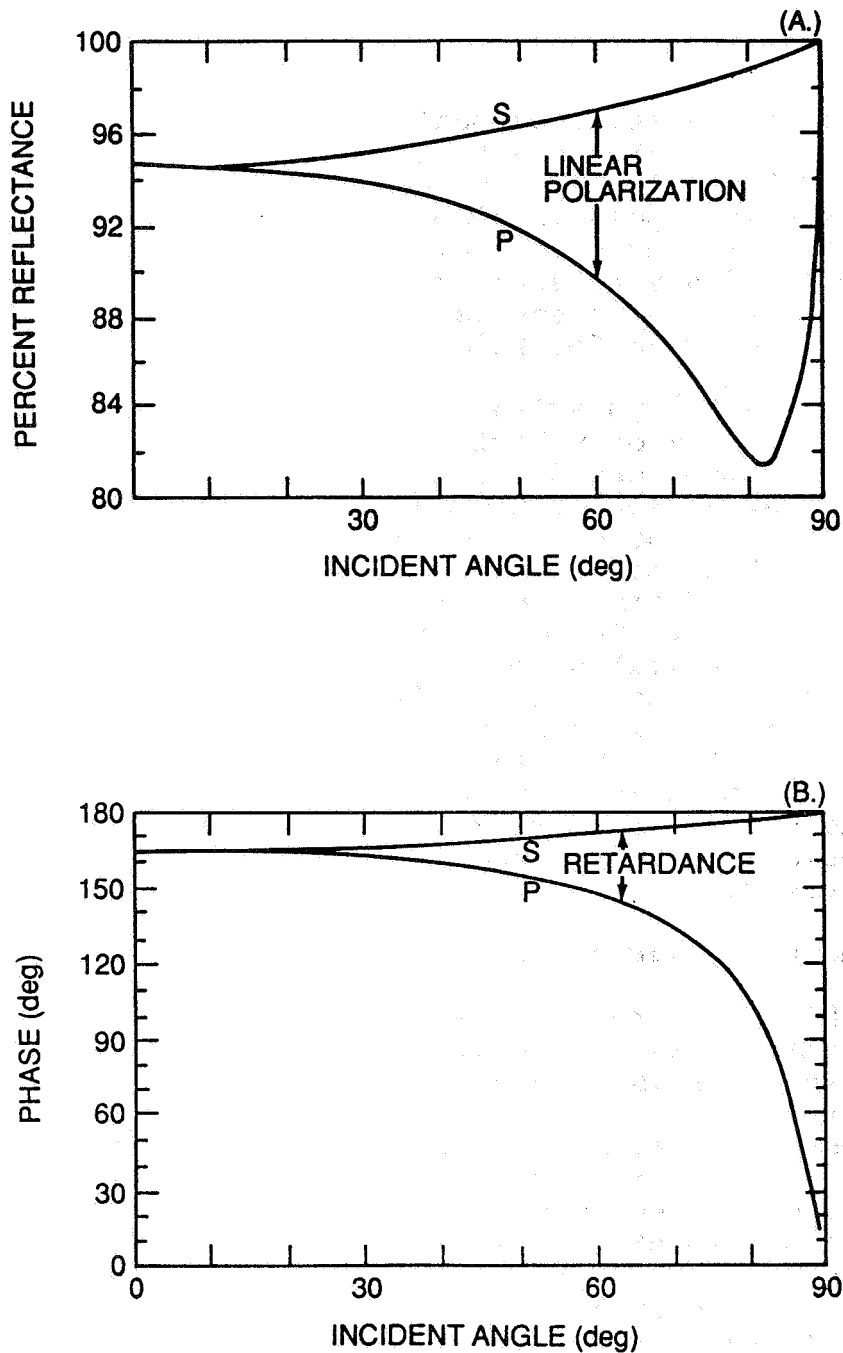


Figure 23. Linear polarization and retardance effects for a Cassegrain telescope. The percent reflectance (a) and phase change (b) on reflection from an aluminum thin-film coating is shown for both the s and p components of the incident light wave. The wavelength of the light is 5250 Å and the complex refractive index of the aluminum coating is $n = 0.7 - 7.0i$. The differences in the s and p reflectance cause linear polarization aligned with the incident plane. The differences in the s and p phase cause a linear retardance. The differences are small for small angles of incidence, but they are not negligible for the SAMEX magnetograph design.

$$R_p = 0.94 - 0.02i^2,$$

where the reflectance rate of change per unit angle of incidence squared (i^2) is approximated from Figure 23a. Therefore, the induced linear polarization is estimated to be

$$L_p = \frac{R_s - R_p}{R_s + R_p} \approx 0.02i^2.$$

At the edge of the primary mirror of aperture D and focal length f_1 , a ray is reflected through a total angle of $\tan(D/f_1) \sim D/f_1 = 1/f$, where f is the f -ratio. Then for a Cassegrain telescope the maximum angle of incidence i_m is given approximately by one half the reciprocal of the f -ratio. For an f -ratio of 5, the angle of incidence is $i_m \approx 1/10$. Hence the induced linear polarization is on the order of $\frac{0.02}{100} \approx 2 \times 10^{-4}$. For a linearly polarized ray this represents a rotation of its plane of polarization over a segment of the mirror, and this rotation introduces errors in the deduced polarization state. In Figure 24 the angle of incidence versus the pupil coordinate for an "illustrative" Cassegrain telescope is shown for an on-axis and off-axis ray. The point to note is that, for the off-axis ray, the average angle of incidence is not zero, and hence there is a net polarization effect associated with the off-axis rays. Furthermore, even on-axis, we average over the square of the incident angle which gives a net polarization contribution, as shall be discussed below.

For an on-axis beam incident on either of the aluminum-coated Cassegrain telescope mirrors, the magnitude and orientation of the linear polarization and linear retardance associated with the mirrors have the forms shown in Figure 25a (polarization) and Figure 25b (retardance) as a function of the pupil coordinates. In Figure 25a, the linear polarization is zero in the center of the mirror where the beam is at normal incidence. The magnitude of the linear polarization (polarizance) increases quadratically with distance from the center of the mirror. The linear polarization is oriented radially. This polarization pattern (polarization aberration) is called linear polarization defocus (or quadpol). Figures 26a and 26b show the effect of large amounts of linear polarization defocus (about 40 times more than the conventional Cassegrain) on beams of uniform, left-circularly polarized light (a) and uniform, vertical-linearly polarized light (b). The linear polarization associated with the telescope mirrors changes the polarization state of the light causing spatial variations of intensity and polarization across the beam. If the polarization state of the light is now measured with a polarimeter, a polarization state different from that incident on the mirrors is obtained. Despite the symmetry associated with the resulting transmitted polarization patterns (such as in Figure 26), the polarization variations do not cancel (due to averaging over the incident angle squared). This is best understood by considering

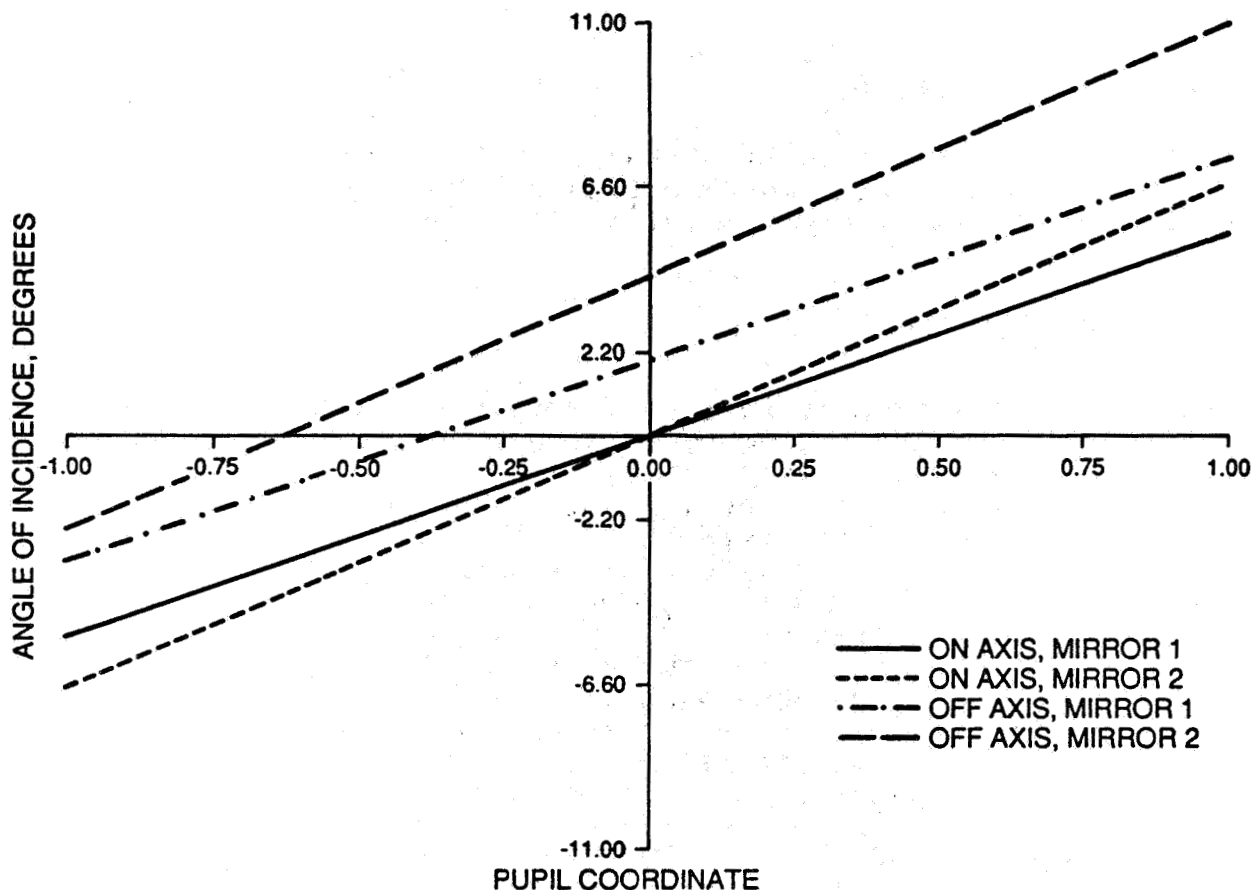
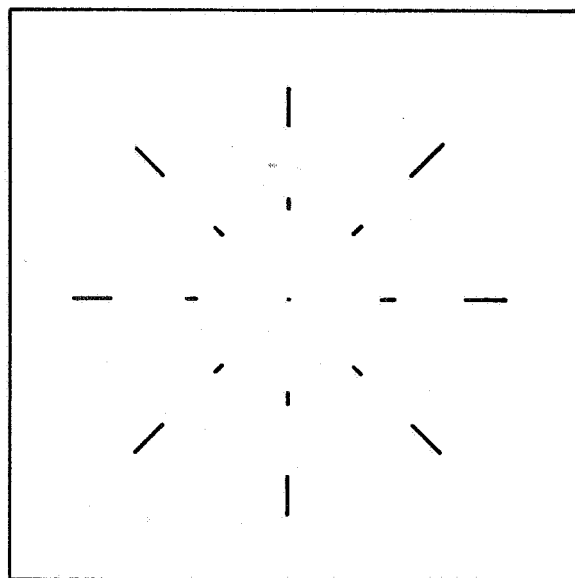
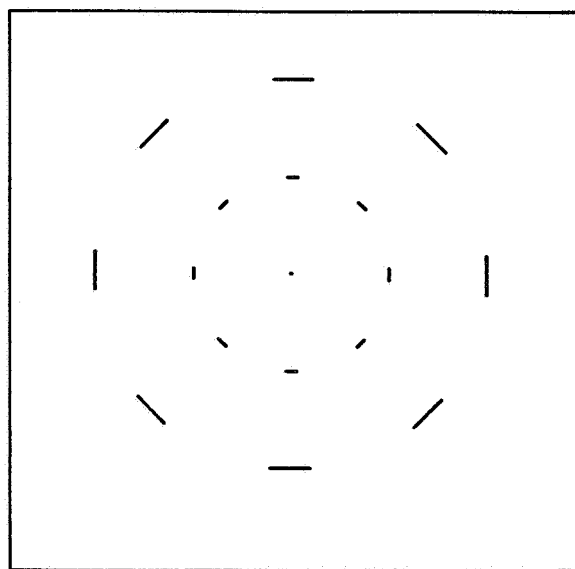


Figure 24. Illustrative example of angles of incidence at Cassegrain mirrors. These curves show the variations of the angle of incidence along the normalized pupil coordinate for on-axis and off-axis rays incident on both the primary and secondary mirrors of a Cassegrain system. The off-axis rays have a non-zero average angle of incidence at both the primary and secondary mirrors. This implies a net linear polarization associated with the telescope for off-axis rays. The example shown here is for a Cassegrain with a larger field of view than the one chosen in the SAMEX design. It was chosen to provide a vivid example of the off-axis problem.



A.



B.

Figure 25. Patterns of linear polarization and retardance for a Cassegrain telescope. The linear polarization pattern (a) and linear retardance pattern (b) associated with an aluminum-coated Cassegrain primary mirror are shown for an on-axis beam. In (a) the magnitude and orientation of the linear polarization are given as linear segments. In (b) the linear segments represent the magnitude and fast-axis orientation of the retardance. Both patterns are quadratic functions of the pupil coordinate and are termed "polarization defocus."

the transmitted light as being a superposition of two polarization states: the incident state yields the correct polarization measurement; the light in the orthogonal state constitutes the error signal introduced by the instrumental polarization.

For example, in the case of Figure 26b, the light in the orthogonal state (horizontal linear polarization) has the form across the pupil shown in Figure 27. Although the phases of this horizontal component are 180° out of phase in the four quadrants (resembling astigmatism), all this light will pass through a horizontal linear polarizer and contribute to the polarization error signal. The phase differences do not cause cancellation of the polarization aberration; instead they affect the polarization accuracy and change the structure of the diffraction pattern (Kubota and Inoue, 1959).

For a metallic mirror, the effects of the linear retardance are orders of magnitude larger than the effects of the weak linear polarization. Figure 25b shows the form of the linear retardance associated with an on-axis beam incident at a Cassegrain telescope mirror. The orientations of the lines represent the orientation of the fast axis of the retardance. The lengths of the lines signify the magnitude of linear retardance, which increases quadratically with pupil coordinate. This polarization aberration is called linear retardance defocus or quadtard. Figures 28a and 28b show the effect of quadtard on a uniform, left-circularly polarized beam (a) and on a uniform, vertical-linearly polarized beam (b). The dominant effect of the retardance is a coupling of linear into circular polarized light and vice versa. These figures are for retardances about 40 times greater than the retardance of a conventional Cassegrain.

In general, a Cassegrain telescope displays both these polarization aberrations simultaneously, linear polarization defocus, and linear retardance defocus, with the retardance being the larger term. This polarization aberration induces polarization coupling which reduces the accuracy of subsequent polarimetric measurements. The polarization coupling for a SAMEX-type Cassegrain design but with standard aluminum coatings is 3×10^{-3} , almost 2 orders of magnitude greater than the radiometric accuracy of the optics. Similar amounts of polarization accuracy would be associated with relay optics utilizing standard antireflection coatings.

3.3 The SAMEX Design

It is clear from the discussion of the previous section that standard optical designs will not suffice for the SAMEX foreoptics because they produce unacceptable levels of instrumental polarization. The design of the SAMEX foreoptics resulted from the development of a new method for the analysis of instrumental polarization based on the theory of "polarization aberrations." This theory allows the description of the

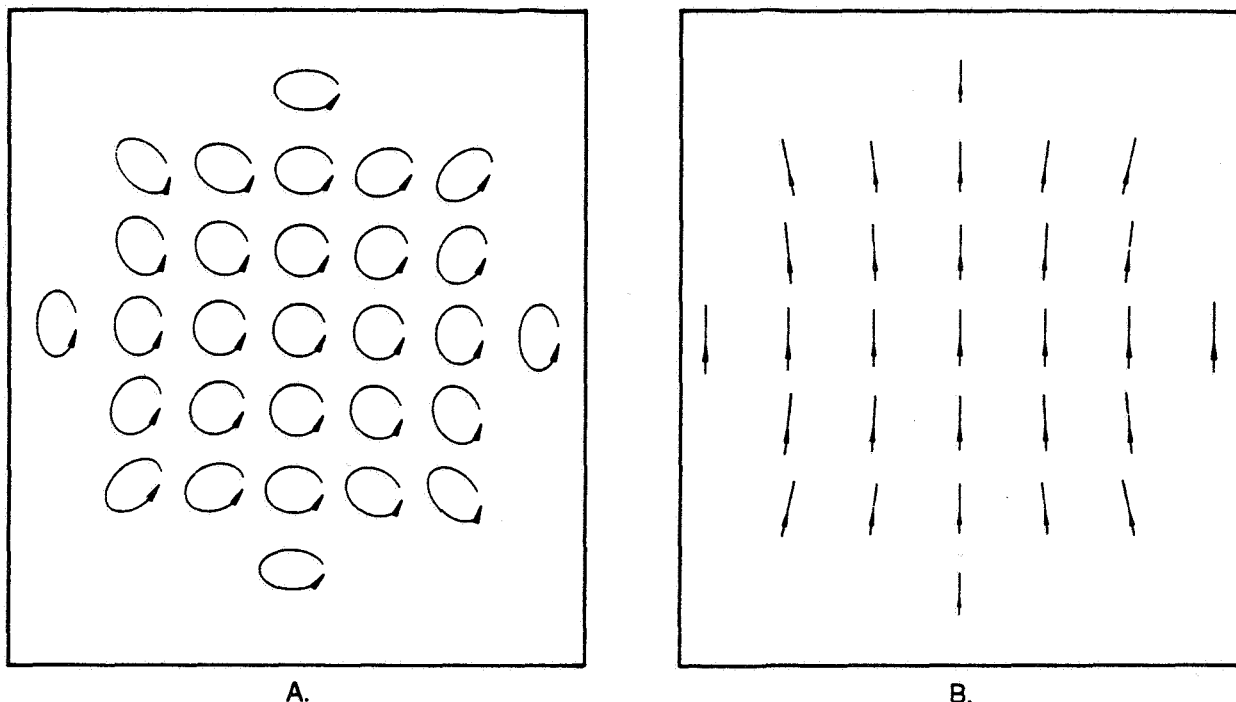


Figure 26. Linear polarization defocus effects for a Cassegrain telescope. This figure illustrates linear polarization defocus effects for a Cassegrain primary mirror for a beam of (a) uniform left circular polarization and of (b) uniform vertical linear polarization. The ellipses and arrows at the top, bottom, left, and right of the diagram represent the same relative position of the polarization ellipse of the reflected beam of a Cassegrain mirror with an exceptionally large field of view (40X the normal Cassegrain field). The location of the arrowheads represents the phase of the light where one cycle is a full wave. The defocus effect introduces the orthogonal state of polarization which represents an error signal introduced by the instrumental polarization in the measurements made with the magnetograph.

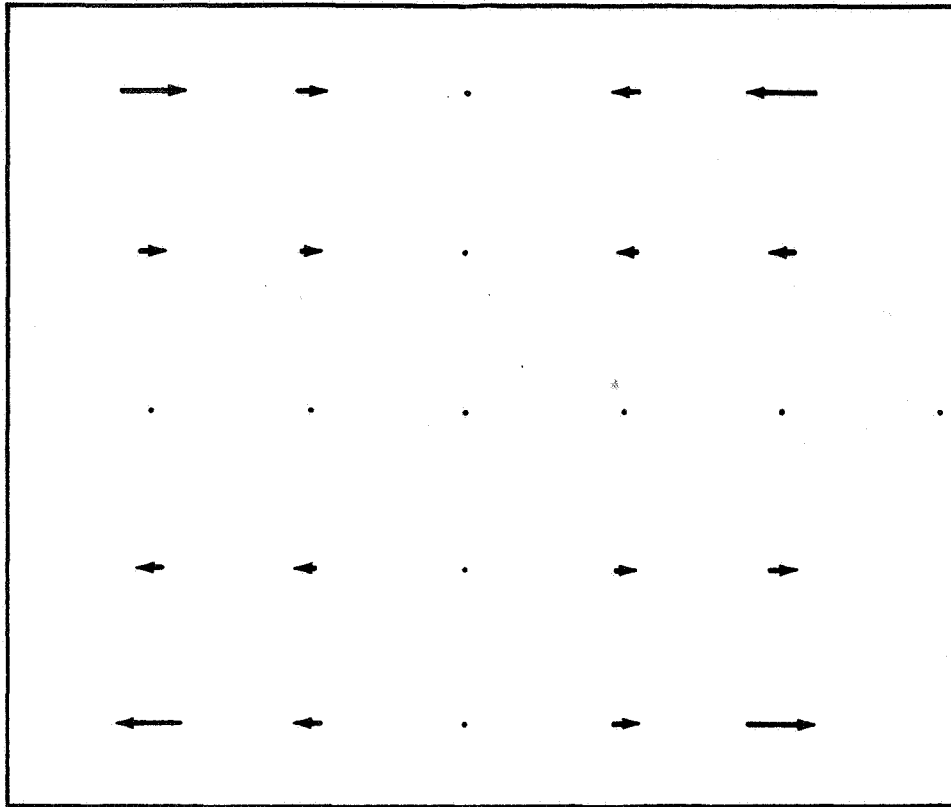
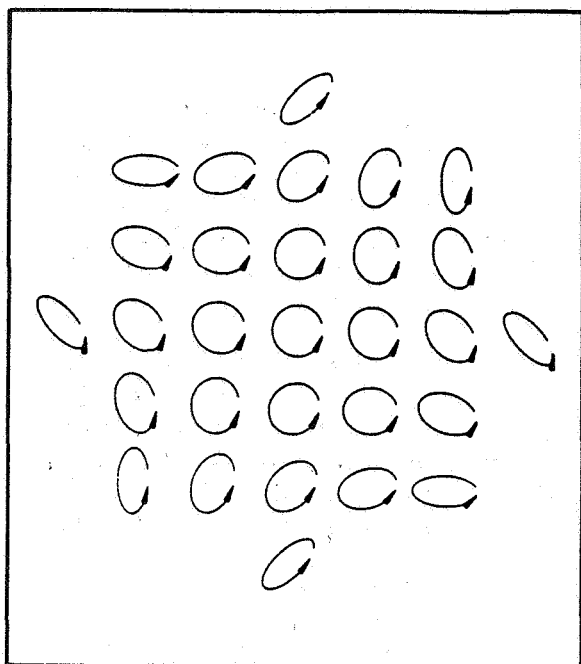
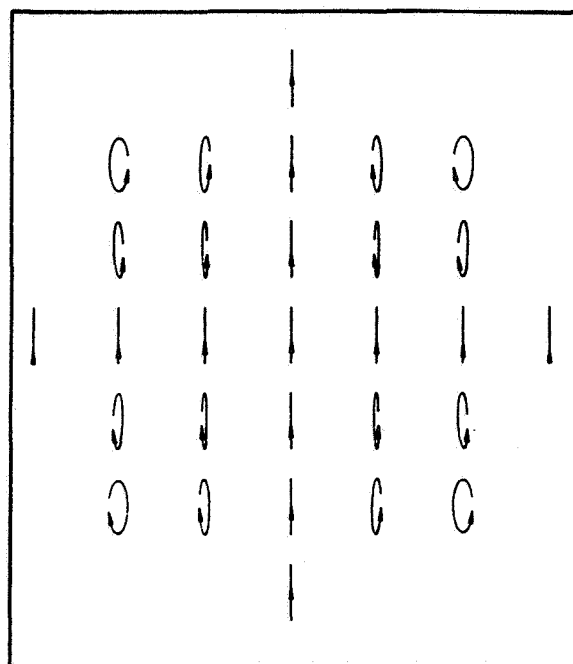


Figure 27. Polarization errors associated with a Cassegrain telescope. This figure illustrates the orthogonal state of polarization introduced by the instrumental polarization of a Cassegrain mirror for the incident linear (vertical) polarization of Figure 26b. These horizontal components will pass through the linear polarizer of a magnetograph's polarimeter when it is in the horizontal position and thereby contribute to the error signal of the magnetograph.



A.



B.

Figure 28. Linear retardance effects associated with a Cassegrain telescope. The large effect of linear retardance from a Cassegrain mirror is shown for (a) a beam of uniform left circular polarization and for (b) a beam of uniform linear (vertical) polarization. The dominant effect of linear retardance is a coupling of linear polarization into circular and vice versa.

variations of amplitude, phase, and polarization of an optical wave front across the exit pupil of an optical system. Because the theory naturally incorporates the polarization properties of the thin-film coatings on the individual surfaces of the optical system, this method integrates the coating design with the lens design. In ordinary optical design work, these two phases of the design are normally decoupled and pursued separately.

This unified optical and coating design was performed for the SAMEX foreoptics to insure that the SAMEX magnetograph will accurately measure the polarization state of incident sunlight. To improve the SAMEX polarization accuracy over standard designs, various telescope and lens coatings were investigated. Most standard reflection-enhancing coatings for telescope mirrors were found to be substantially worse than the bare aluminum. Two coatings were designed which had significantly improved polarization performance for a telescope. These coatings are described in the section on coatings (section 3.6). Figure 29 shows the linear polarization (a) and linear retardance (b) associated with a Cassegrain telescope when these coatings are utilized and when ordinary aluminum coatings are used, as a function of the height of the ray entering the telescope, from top to bottom. The amount of polarization or retardance for the three coating choices is plotted along the y-axis for (1) two aluminum-coated mirrors, (2) two Q201-coated mirrors, and (3) one Q201-coated mirror and one eight-layer, enhanced-reflection coating. The third design (3) balances polarization and retardance effects of opposite signs between the two mirrors to achieve a polarization performance superior to either mirror separately.

Similar design strategies have been used with the relay lenses to achieve significant improvements over conventional coatings and to balance the remaining polarization effects. As a result, the overall polarization performance for the foreoptics has been improved by a factor of 10,000 relative to conventional designs. Thus the actual attainment of the polarimetric sensitivity determined by the SAMEX requirements will presumably be limited only by whatever scattering effects are present, and not by the instrumental polarization of the optics.

The method used to achieve this dramatic result - polarization aberration theory - is outlined briefly in sections 3.4 and 3.5 and expanded upon in Appendix C. A full development of the method is given in Chipman (1987). The basic results from the method are the second-order polarization aberration coefficients which provide a quantitative measure of the polarization accuracy of any optical system. Specifically, these coefficients determine the parameter Δ_p , the polarization accuracy, defined as the maximum fraction of light which can be coupled into an orthogonal polarization state. It is given in terms of the second-order polarization aberration coefficients $P(1,0,2,2)$, $P(1,1,1,1)$, and $P(1,2,0,0)$:

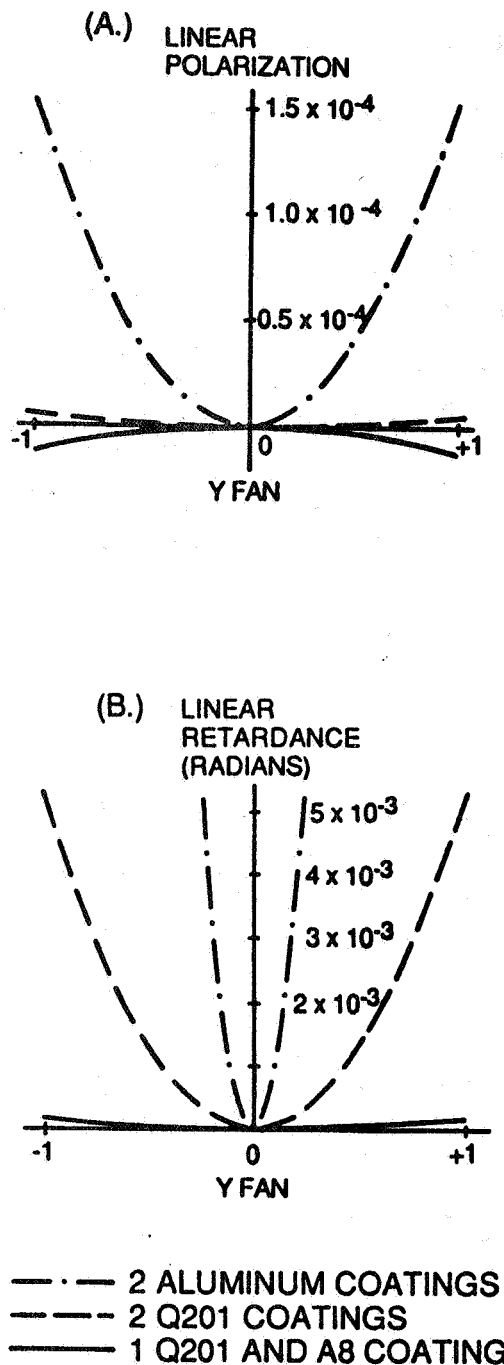


Figure 29. Instrumental polarization effects for rays through a Cassegrain telescope. Linear polarization (a) and retardance (b) as a function of pupil coordinate in a Cassegrain telescope are plotted for three different sets of reflective coatings on the primary and secondary mirrors. The three sets of coatings are: (1) two aluminum (dot-dash curve), (2) two Q201 coatings (dashed curve), and (3) one Q201 and one aluminum (solid curve). The Q201 interference coating is a 201 coating of bi-layers with high ($n = 1.52$) and low ($n = 1.45$) refractive indices.

$$\Delta_p < P(1,2,0,0)^2 + \frac{1}{2} P(1,1,1,1)^2 + \frac{1}{3} P(1,0,2,2)^2 .$$

For the SAMEX magnetograph foreoptics and coating design given herein, the value of $\Delta_p < 1.4 \times 10^{-3}$ was obtained. For standard coatings of aluminum, the polarization accuracy is $\Delta_p = 2.7 \times 10^{-3}$ for a Cassegrain telescope alone (no relay lenses). (The second-order aberration coefficients used in these calculations are given in Table 19a.) This result for Δ_p means we have achieved our design goal: the polarization state of the light from the Sun can be determined without introducing polarizing effects from optical elements in the system. We have been able to effectively eliminate the problem of induced instrumental polarization in the SAMEX foreoptics.

3.4 Polarization Aberration Theory

The SAMEX foreoptics are intended to transmit all polarization states equally. But all optical interfaces display some polarization when used at non-normal incidence. Thus polarization is present in all systems at some level. If the system is intended to be non-polarizing, the instrumental polarization is often termed "residual polarization" to signify its generally undesirable character. Residual polarization might be compared to wave front aberration because both interfere with the measurement of optical fields and reduce the image forming potential of the optical system.

The principal cause of instrumental polarization in the optical systems of present solar magnetographs is the polarization due to non-normal incidence at the optical interfaces and coatings. Since each ray takes a different path through the system with its own angles of incidence and planes of incidence, each ray in general experiences a different change in its state of polarization. This residual polarization varies with wavelength, object coordinates and pupil coordinates. "Polarization aberrations" will be defined as variations of the amplitude, phase and polarization of an optical wave front across the exit pupil of an optical system and the dependence of these variations on wavelength and object coordinate. The polarization aberrations are extensions of the wave front aberrations and encompass both amplitude and polarization variations, thus providing a more complete characterization of the electromagnetic fields transmitted by an optical system.

Vacuum-deposited thin films are used on most optical surfaces to control the amount of transmission and reflection. These thin films are usually less than the wavelength of light in thickness. Being so very thin, the effect of the films on ray paths are accurately modeled by treating the films as having parallel surfaces which contour the substrates on which they are deposited. Due to the closely spaced parallel surfaces, thin films have negligible influence on the ray paths through the

system and are generally ignored when simulating a system by ray tracing. These coatings principally affect the amplitude and polarization of the ray and have much less effect on the optical path difference. This division, with the optical surfaces governing the ray paths and the thin-film coatings governing the amplitude and transmission, allows the optical system design problem to neatly decouple into two separate problems, lens design, and coating design. The wave front performance and image quality of the system are calculated by a lens designer using a ray tracing optical design program. The amplitude and polarization calculations at individual surfaces are performed using a thin-film design program.

This decoupling of optical design and coating design has usually worked well. The coatings designed to optimize the transmittance or reflectance at an interface have usually reduced the amplitude and polarization variations and thus reduced the polarization aberrations at the interface as well. For example, a quarter-wave magnesium fluoride antireflection coating on glass typically reduces reflection losses at the design wavelength by a factor of 4, and reduces the instrumental polarization by a comparable factor. This fortuitous circumstance has allowed lens and coating design to remain relatively uncoupled. Thus instrumental polarization was usually ignored as a higher order effect. But it is not sufficient to design thin-film coatings in isolation from the lens design for the SAMEX magnetograph - the demands on amplitude and polarization performance are too great.

For the designs of the SAMEX system special methods have been developed to calculate the instrumental polarization of the SAMEX foreoptics. These methods are described in detail in Appendix C. Calculating the instrumental polarization requires performing thin-film calculations during the ray tracing process. This idea is not new, but its implementation is complex enough to have delayed this obvious integration of these two branches of optical design until specifications required it.

In this new methodology, a Jones polarization matrix is calculated for arbitrary optical paths through the optical system and includes the effects of all the specified optical coatings on the curved optical interfaces. The specific technique used for this calculation is the method of "polarization aberrations." This is a method analogous to the aberrations of geometrical optics (spherical aberration, coma, astigmatism, etc.) except that it encompasses the polarization effects of coatings as well as the wave front aberrations.

The analysis of the SAMEX optics proceeds in several stages and is summarized in Table 16. First, the optical system is designed using the CODE V optical design program to optimize the optical design for a spatial resolution with a minimum aperture. During this phase the angles of incidence are kept as small as possible to reduce polarization effects from coatings. Second, special thin-film, reflection-enhancing coatings are

designed for the telescope mirrors and special antireflection coatings are designed for the lenses because conventional coating designs show significant polarization and retardance contributions near normal incidence. These coatings are special designs which reduce the polarization effects at small angles of incidence over the wavelength band of the magnetograph. Next, a Taylor series is calculated to represent the coating performance. Finally, the effect of these coatings in the magnetograph optical design is calculated using the polarization aberration method; this last step produces the parameter Δ_p , the polarization accuracy.

Table 16. Steps in instrumental polarization simulation

-
1. Design optical system using CODE V.
 2. Design low polarization, high reflectivity telescope coatings.
 3. Design low polarization, antireflective lens coatings.
 4. Determine Taylor series representation of coating performance.
 5. Calculate polarization aberration coefficients of optical system.
 6. Calculate Δ_p .
 7. Iterate 2, 3, 4, 5, 6 until satisfactory performance is achieved ($\Delta_p < 10^{-5}$).
-

In the following section, the method for deriving the polarization aberration coefficients and the polarization accuracy Δ_p is outlined in more detail.

3.5 Polarization Aberration Coefficients and Polarization Accuracy

The polarization states of the electromagnetic field are described by the complex two-component Jones vector, \vec{J} . The polarization states and hence the Jones vector are transformed when a ray passes through an optical interface which is described by the Jones matrix transformation JJ . For the SAMEX foreoptics we have homogeneous, weakly polarizing optical elements (by design) for which the transmission coefficients perpendicular, t_s , and parallel, t_p , are given in terms of a Taylor series expansion of the angles of the chief and marginal rays, i_c and

i_m , respectively. Each individual optical ray path for the rotationally symmetric system can be defined in terms of the object height, H , and the pupil coordinates, ρ and ϕ as defined in Figure 30. The cascade effect of an optical train is given by the product of the individual Jones transformation matrices for each element interface. Then the overall polarization of the foreoptics is described by the complex Jones transformation matrix which is expanded in terms of the ray coordinates (ρ , ϕ , and H) and the basis matrix set $\sigma(k)$:

$$JJ = \sum_{k=0}^3 \sum_u \sum_v \sum_w P(k,u,v,w) H^u \rho^v \cos^w(\phi) \sigma(k)$$

where $P(k,u,v,w)$ are the expansion coefficients. The 2 by 2 matrices, $\sigma(k)$, which define the k^{th} polarization state, are the identity matrix and the Pauli spin matrices. This expansion is in the same form as the standard wave front aberration expansion. These polarization expansion coefficients for an expansion to second order (specified by $s = 2$) in the angles of incidence are a function of the total optical transmission, τ , the normalized second-order basis set coefficients, $d(k,s,q)$, for each individual optical element, q , and the angle of incidence of the chief and marginal rays. Hence these coefficients are dependent on the characteristics of the optical coatings and the ray tracing results. These SAMEX specifications are given in Tables 17-19.

The polarization aberration expansion for the radially symmetric system of interfaces with isotropic coatings and non-polarizing transparent media is given to second order in the angles of incidence by the polarization aberration expansion

$$JJ(H, \rho, \phi) = \tau [\sigma_0 + P(1,2,0,0)H^2\sigma_1 + \\ P(1,1,1,1)H\rho(\sigma_1\cos\phi - \sigma_2\sin\phi) + \\ P(1,0,2,2)\rho^2(\sigma_1\cos 2\phi - \sigma_2\sin 2\phi)].$$

The amplitude a_{ijk1} of the polarization aberration coefficient $P(i,j,k,1)$ describes the magnitude of the polarization aberration effects while the phase δ_{ijk1} of the coefficient describes the magnitude of the retardances. Then, given a specific incident polarization state, the polarization state of the exit beam can be calculated from this matrix. Therefore the amount of polarization in the orthogonal state can be determined. This orthogonal polarization in the exit beam then determines the accuracy Δ_p of the polarization measurement.

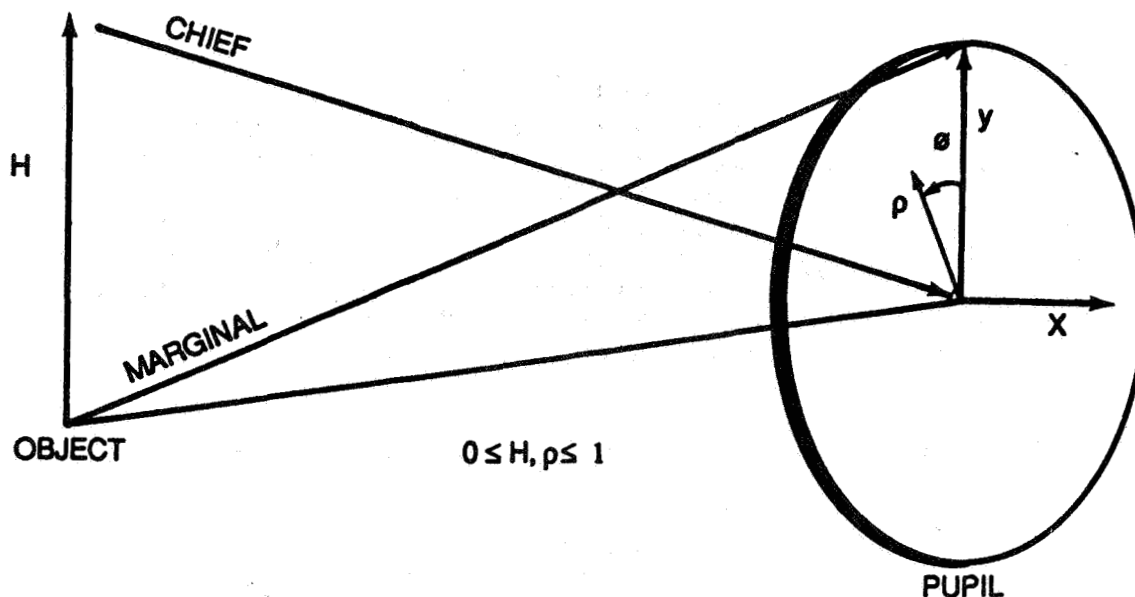


Figure 30. Paraxial coordinate system. The paraxial system is a normalized right-handed coordinate system. The z axis is the optical axis of a rotationally symmetric optical system; light initially travels in the direction of increasing z . Rays through an optical system are characterized by ray coordinates at the object and entrance pupil. H is the normalized object coordinate, ρ is the normalized pupil radius, and ϕ is the polar angle in the pupil measured counterclockwise from the y axis. The normalized Cartesian coordinates in the pupil are x and y . The chief and marginal rays are also shown.

As an example of the coupling of the optics to an orthogonal state of polarization, i.e., polarization cross talk, consider the following example. The on-axis linear polarization and linear retardance of the SAMEX foreoptics, i.e., the term linear defocus, is described by the fourth term in the polarization aberration expansion JJ. The instrumental polarization function $J_d(H, \rho, \phi)$ for linear defocus is then

$$\begin{aligned} JJ_d(H, \rho, \phi) &= \tau [\sigma_0 + P(1, 0, 2, 2) \rho^2 (\sigma_1 \cos 2\phi - \sigma_2 \sin 2\phi)] \\ &= \tau [\sigma_0 + (a_{1022} + \tilde{\delta}_{1022}) \rho^2 (\sigma_1 \cos 2\phi - \sigma_2 \sin 2\phi)], \end{aligned}$$

where τ is the amplitude of transmittance of the system down the optical axis. The transmittance (τ) describes the polarization independent reflection and absorption losses associated with the ray down the optical axis. $P(1, 0, 2, 2)$ describes the linear polarization (a_{1022}) and linear retardance (δ_{1022}) associated with the marginal ray.

At any given point in the pupil, the eigenpolarizations are linearly polarized light, oriented radially, J_r , and tangentially J_t . This concept is important to the definition of polarization accuracy which is to be defined. We shall now illustrate the effect of crosstalk for incident circularly and linearly polarized light.

Maximum coupling occurs when the incident light is circularly polarized, since circularly polarized light can always be decomposed into equal components of J_r and J_t everywhere in the pupil. The coupling is zero in the center of the pupil (where the polarization and retardance vanish) and increases to a maximum coupling at $\rho = 1$ of $I_C^1(H, 1, \phi) = |P(1, 0, 2, 2)|^2 = a_{1022}^2 + \delta_{1022}^2$ at the edge of the pupil. The net fraction of incident circularly polarized light coupled into the orthogonal circularly polarized state, I_C^1 , is given by the integral of circularly polarized intensity over the pupil:

$$\begin{aligned} I_C^1 &= \frac{1}{\pi} \int_0^{2\pi} d\phi \int_0^1 \rho d\rho |\tau \rho^2 P(1, 0, 2, 2)|^2, \\ &= \frac{|\tau P(1, 0, 2, 2)|^2}{3}. \end{aligned}$$

If we have a Stokes vector of $+V$ then the amount of polarization in the orthogonal state $-V$ can be calculated. If $-V$ is the maximum amount of cross talk, we identify this as the polarization accuracy.

For incident linearly or elliptically polarized light, the fraction of coupled intensity is less because the light is not composed of equal fractions of eigenstates. The coupling is minimum for incident linearly polarized light, which will be in one of the eigenpolarizations along one axis in the pupil and in the orthogonal eigenpolarization along the orthogonal axis. Here the fraction of coupled energy will be calculated assuming an incident polarization state of horizontal linearly polarized light \hat{H} for calculational simplicity; the same fraction is coupled for any linearly polarized incident state. The orthogonal state of vertical linearly polarized light is designated as \hat{V} . The polarization state transmitted by an optical system described by linear polarization defocus for \hat{H} is

$$\hat{J}(H, \rho, \phi) = \tau[\hat{H} + P(1, 0, 2, 2)\rho^2(\hat{H} \cos 2\phi - \hat{V} \sin 2\phi)].$$

The fraction of incident \hat{H} light coupled into \hat{V} light is equal to

$$\begin{aligned} I_c^2 &= \frac{1}{\pi} \int_0^{2\pi} d\phi \int_0^1 \rho d\rho |\hat{J}(H, \rho, \phi) \cdot \hat{V}|^2 \\ &= \frac{|\tau|^2}{\pi} \int_0^{2\pi} d\phi \int_0^1 \rho d\rho |P(1, 0, 2, 2)\rho^2 \sin 2\phi|^2 \\ &= \frac{|\tau P(1, 0, 2, 2)|^2}{\pi} \int_0^{2\pi} \sin^2 2\phi d\phi \int_0^1 \rho^5 d\rho \\ &= \left| \frac{\tau P(1, 0, 2, 2)}{6} \right|^2. \end{aligned}$$

Therefore, we defined the polarization accuracy, Δ_p , as the maximum fraction of light which can be coupled into orthogonal polarization states. The incident polarized state is given by the Jones vector, \hat{J} . This vector is rotated by the optical system and the rotation is given by the Jones matrix, JJ . The amount of polarization along the orthogonal state of polarization, \hat{J}' , of the incident polarization state is given by the projection of $JJ(\hat{J})$ into \hat{J}' , i.e.,

$$JJ(\hat{J}) \cdot \hat{J}'.$$

The maximum projection is used in defining the polarization accuracy. In the above example $I_1^1 > I_2^2$, and only I_c^1 is retained for the linear defocus term, $P(1, 0, 2, 2)$. This value is given by

the square of the second-order Jones matrix and is given in terms of the polarization aberration coefficients. Then the polarization accuracy is given in terms of the second-order polarization aberration coefficients $P(1,0,2,2)$, $P(1,1,1,1)$, and $P(1,2,0,0)$:

$$\Delta_p = \frac{1}{\pi} \int |JJ_2(1, \rho, \phi)|^2 \rho d\rho d\phi$$

or

$$\Delta_p \leq |P(1,2,0,0)|^2 + \frac{1}{2} |P(1,1,1,1)|^2 + \frac{1}{3} |P(1,0,2,2)|^2,$$

where the integrations have been carried out for the squared terms and estimated for the cross terms as being less than or equal to the direct product terms. For the SAMEX magnetograph design, the value for the polarization accuracy,

$$\Delta_p < 1.4 \times 10^{-7}$$

is obtained for specially selected optical coatings. These second-order aberration coefficients are given in Table 19a. The second-order coefficients are sufficient since the next order that contributes is the fourth order and their polarization effects would be on the order of $(\Delta_p)^2$. In the following section we discuss the process used to design these special optical coatings.

Table 17. Table of total second-order polarization aberration coefficients for the SAMEX magnetograph given the individual surface components of Table 19b.

Vector Quadratic Piston:

$$\begin{aligned} P(1,2,0,0) &= \sum_q d(1,2,q) i_c^2(q) \\ &= -7.72 \times 10^{-5} - i 1.50 \times 10^{-4} \text{ radians}^2 \end{aligned}$$

Vector Tilt:

$$\begin{aligned} P(1,1,1,1) &= 2 \sum_q d(1,2,q) i_c(q) i_m(q) \\ &= -1.42 \times 10^{-5} - i 2.70 \times 10^{-5} \text{ radians}^2 \end{aligned}$$

Vector Defocus:

$$\begin{aligned} P(1,0,2,2) &= \sum_q d(1,2,q) i_m^2(q) \\ &= -1.91 \times 10^{-5} + i 1.80 \times 10^{-4} \text{ radians}^2 \end{aligned}$$

Table 18. Surface by surface polarization aberration contributions given surfaces as defined in Table 19c.

z	Vector Quadratic Piston		Tilt Vector		Vector Defocus	
	$d(1,2) i_c^2(q)$		$2 d(1,2) i_c(q) i_m(q)$		$d(1,2) i_m^2(q)$	
	Re	Im	Re	Im	Re	Im
1	8.3e-10	8.4e-7	-3.9e-8	-3.9e-5	1.8e-6	1.8e-3
2	-1.1e-8	-1.3e-6	3.9e-7	4.6e-5	-1.4e-5	-1.7e-3
3	-4.7e-7	-1.1e-6	4.4e-7	5.8e-7	-1.3e-7	-3.0e-7
4	1.1e-5	-2.6e-5	-3.1e-6	-7.5e-6	-9.1e-7	-2.2e-6
5	-4.9e-6	-1.2e-5	3.6e-7	8.5e-7	-2.6e-8	-6.1e-8
6	-1.3e-6	0	-2.3e-6	0	-3.8e-7	0
7	-5.1e-6	-1.2e-5	-1.4e-6	-3.4e-6	-4.1e-7	-9.8e-7
8	-2.8e-5	-6.6e-5	-2.3e-6	-5.5e-6	-1.9e-7	-4.6e-7
9	-1.3e-5	-3.1e-5	-1.5e-6	-3.6e-6	-1.7e-7	-4.2e-7
10	-1.1e-8	1.1e-6	5.1e-8	-4.9e-6	-2.3e-7	-2.2e-5
11	-5.7e-7	0	1.2e-6	0	-2.7e-6	0
12	-1.1e-6	-2.7e-6	1.3e-6	3.1e-6	-1.5e-6	-3.6e-6

Note: The surface numbers S(i) are related to the surface numbers q by $S(i) = q + 2$, i.e., q = 1 is the primary mirror and the prefilter is not considered. The notation 8.3e-10 means 8.3×10^{-10} .

Table 19. The surface definitions for the polarization aberration coefficient calculations.

Surface q	Optical Coating	Paraxial Angle		Second-Order Polarization**		Magnitude of Second-Order Polarization
		$i_c(q)$	$i_m(q)$	Re	Im	
1	refl201	0.00134	-0.06250	2.82e-7*	2.84e-4	2.8e-4
2	layers8onAl	0.00236	-0.08519	-1.20e-6	-1.39e-4	1.4e-4
3	vl0c5256	0.03862	-0.02031	-1.90e-7	-4.53e-7	4.9e-7
4	"	0.18664	0.05388	"	"	"
5	"	0.12576	-0.00908	"	"	"
6	kf9fk5	-0.30317	-0.05091	-8.91e-8	0.0	8.9e-8
7	vl0c5256	-0.12743	-0.03622	-1.90e-7	-4.53e-7	4.9e-7
8	"	-0.29808	-0.02495	"	"	"
9	"	-0.20326	-0.2363	"	"	"
10	vl0c525	-0.00703	0.03169	-1.38e-7	1.33e-5	1.3e-5
11	kf9fk5	-0.06243	0.13588	-8.91e-8	0.0	8.9e-8
12	vc10c5256	-0.5980	0.06974	-1.90e-7	-4.53e-7	4.9e-7

*(2.82e-7 = 2.82×10^{-7})

**The rate of linear polarization (Re) and retardance (Im) per angle of incidence squared (deg^{-2}) for the specified optical coating.

3.6 Coating Designs

A design study was performed to find coatings especially suitable for use in the SAMEX magnetograph. First, many conventional coatings were examined and found to have polarization and retardance properties that make them unacceptable for use in the SAMEX magnetograph. Alternative coating designs were devised with improved polarization performance and the best of these were specified for the instrumental polarization simulation. Coatings were designed, optimized, and analyzed using two thin-film coating programs: FTG Software's Filmstar and Prof. A. Macleod's (University of Arizona) FILMS.

The requirements for the SAMEX magnetograph coatings are the following:

1. They must have extremely low polarization properties near normal incidence so that they introduce a minimum of polarization or retardation into the optical system. Specifically, the differences between the s and p amplitude coefficients for reflection or transmission for less than 10° angle of incidence should be 0.01% and the phase difference should be less than 0.05° .
2. The system's transmission must not be compromised by the coatings: reflective coatings should reflect >99.9% and transmission coatings transmit >99.7%.
3. They must be manufacturable and not be excessively costly.
4. The amount of scattered light should be small.
5. They should have reasonable tolerances for fabrication.

Several common reflecting coatings for telescopes and antireflecting coatings for lenses were analyzed and found to be inadequate for the SAMEX magnetograph. For example, telescopes are frequently coated with bare aluminum coatings (Al) or quarter-wave stacks on top of aluminum. These coatings have very little linear polarization near normal incidence but substantial amounts of linear retardance. The common antireflection coatings for lenses (quarter-wave MgF_2 and broadband antireflective coatings) have very little retardance near normal incidence but substantial linear polarization.

Three coatings (see Table 20) have been selected from our coating design study for inclusion in the SAMEX magnetograph design. The polarization effect of these coatings has been used in the instrumental polarization simulation. These coatings demonstrate that coatings with the necessary low polarization performance can be designed. The present coatings are somewhat

sensitive to fabrication errors and are very wavelength sensitive. The designs would benefit from further work to understand why they work, knowledge which should lead to fully optimized designs with improved manufacturing tolerances.

Table 20 contains the description of the three coatings specified. In addition, the optical system contains two cemented interfaces, between kf9 glass and fk5 glass, specified kf9fk5. The polarization of this interface, although negligible, was included in the instrumental polarization calculation.

The coating Q201 is a 201-layer, quarter-wave enhanced reflection coating specified for the primary mirror. The philosophy behind this design is that the polarization effects arise because of large refractive index differences between layers. This coating uses two materials with a small index difference, fused silica at $n = 1.45$ and evaporated glass at $n = 1.52$. To achieve the desired reflection performance with smaller index differences requires more layers, 201 in this case. The two materials chosen are both amorphous and thus quite suitable for coatings with large numbers of layers, since they don't display the microcrystalline growth patterns which lead to the unacceptable scattering and inhomogeneity associated with crystalline materials. Figure 31 shows performance curves for Q201 both as a function of angle of incidence and wavelength. Panel A1 of Figure 31 shows the absolute phase change on reflection for the s and p components. The quadratic portion of these curves is the defocus introduced by the coating. The parameter of greatest interest is the retardance, the difference between the s and p phase changes; it is plotted separately in Panel A2. This retardance is below 0.05° for the angles of incidence less than 5° . Panel A3 shows the s and p intensity reflectance as a function of angle of incidence. The difference between the s and p reflectance is the linear polarization, here less than 0.01%.

The coating specified for the secondary, L8onA1, is a more conventional 8-layer, reflection-enhancing coating specially optimized to complement the Q201 coating. This coating has the opposite sign on both its linear polarization and retardance relative to Q201. Thus the small residual polarizations of these two coatings tend to cancel, resulting in a nearly polarization free Cassegrain design.

The coatings specified for the lens surfaces are a 2-layer coating of the V coating family of designs. A region of solution was found with the remarkable property that the linear polarization and linear retardance both changed signs within 50 nm in wavelength of each other. Thus, by varying the thicknesses of the layers, low polarization coatings with several different useful properties were obtained. The two scalings used in the design were 1.0 for the coating vl0c525 and 1.06 for vl05256. These coatings have by far the lowest polarization effects of any lens coatings investigated. Fabricating and testing samples of these coatings should be conducted in the next study phase.

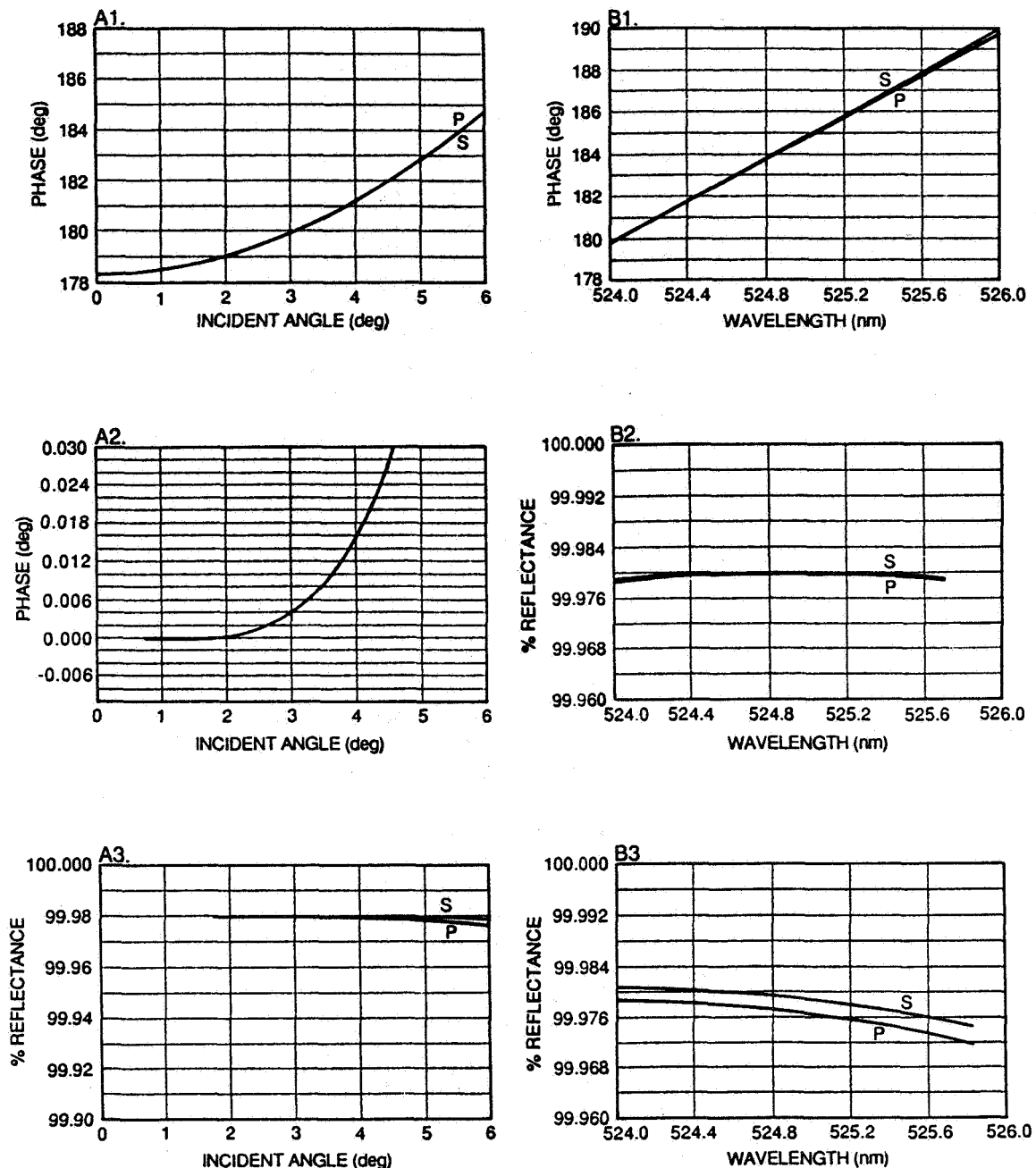


Figure 31. Optical properties of the interference coating Q201. The phase and reflectance properties of the interference coating Q201 designed for the SAMEX telescope mirrors are shown as functions of the angle of incidence (A, on the left) and wavelength (B, on the right). (A1) The phase of a reflected beam is shown as a function of the angle of incidence. The s and p components of an incident wave of wavelength 5250 Å are shown as one curve. (A2) The difference in phase of the s and p components shown in (A1) is plotted as a function of the angle of incidence. (A3) The percent reflectance of a 5250 Å-reflected wave is shown as a function of the angle of incidence. (B1) The phase of the reflected wave is plotted as a function of wavelength. (B2) The percent reflectance of a wave incident at an angle of 2° is shown as a function of wavelength. (B3) The percent reflectance of a wave incident at an angle of 6° is shown as a function of wavelength.

Table 20. Coatings specified for the magnetograph.

Name	Type	Number of Layers	Polarization Magnitude	Reflectance or Transmission	Coating Design
V10C525	Antireflection	2	4.9e-7	99.7%	(0.25H 0.25L)100 0.254H H=1.52 L=1.45
Q201	Reflective	201	2.8e-4	99.98%	(0.062H 0.82L) H=2.15 L=1.38
L8onAl	Reflective	9	1.4e-4	99.65%	0.38H 0.29L 0.26H 0.25L (0.25H 0.25L)2 H=2.38 L=1.38

All the polarization analyses of thin films performed here have assumed ideal, thin-film structures, materials which are uniform, isotropic, homogeneous, and free from scattering. This is a sufficient assumption for the design and analysis of the thin films. Actual thin films possess extremely complex microstructure which contributes to scattering, inhomogeneity, and anisotropy. The typical thin film microstructure is an array of columns growing up out of the substrate. This causes non-ideal coating properties which will impair the performance of the system at some level.

The performance of the SAMEX magnetograph will be enhanced by using the finest coatings available, which will probably be deposited using ion-assisted, thin-film deposition. By bombarding the growing thin film with energetic ions, usually helium or argon, the growing thin-film structure is pressed down and becomes denser (P. J. Martin, 1986). This disrupts the growth of the large columnar structures which are responsible for much of the scattering and anisotropy. The resulting films have lower scatter and greater uniformity. The ion-assisted, thin-film deposition technique is still new and not yet widespread in its use.

As part of the preliminary efforts leading to the actual fabrication of the SAMEX magnetograph, test coatings need to be prepared and coating vendors qualified. Several coating manufacturers should be contracted to produce small (2 x 2 inch) samples of the specified coatings. These coatings should be tested for spectral transmission, polarization performance and scatter (bidirectional reflection measurements) as part of the final design and vendor qualification. The Marshall Space Flight Center and The University of Alabama in Huntsville Center for Applied Optics have the expertise for such coating characterization. These data would then be used to calculate the

impact of the stray light coming from the coating scatter on the magnetic field accuracy of the magnetograph.

3.7 Further Sources of Instrumental Polarization

Two effects that could degrade the polarization performance of the SAMEX optics are coating anisotropy and scattered light. Coating anisotropy is the variation of refractive index with direction in the coating. Coating anisotropy can be measured ellipsometrically. It frequently occurs where coatings have been deposited at non-normal incidence. Most coatings have a columnar microstructure which usually grows out of the substrate toward the source. For a non-normal deposition angle, the resulting coatings have built-in birefringence. For the magnetograph optical design, this additional coating-induced retardance is a problem which must be held to acceptable levels in the final coatings. A level of anisotropy below a root sum squared birefringence of 0.0001 radians rms per coating is highly desirable.

Scattering depolarizes light: the scattered light is random and carries less information about its original polarization state. Coatings, because of their detailed microstructure, may display substantial scattering. Considerable effort is being devoted on many approaches to producing optics with reduced scatter, including the ion-assisted deposition coatings already mentioned. Scattering in the coatings of the foreoptics will probably be the limiting factor in the accuracy of the magnetograph, now that the instrumental polarization of the Cassegrain telescope has been reduced by orders of magnitude.

3.8 Optical Tolerances

An optical tolerance study of the magnetograph is outside the scope of this study. However we would like to point out the areas of special study that are needed for a flight instrument.

The individual optical components need to be qualified for spaceflight. This would include operation under a vacuum for an extended period of time. The effects of particle radiation over the lifetime of the spacecraft must be resolved. The mechanical robustness of the optical system is to be defined. This would include the effect of thermal drift on the optical alignment. The general effects of mechanical fitting and sensitivity due to fabrication errors need to be addressed, including surface figure, tilts, decentrations, and optical coating variations. The overall alignment and calibration procedures are to be specified before the optical design is completed (Yoder, 1986).

4. The Polarimeter

4.1 Introduction

We have determined that a flat rotating polarizer, a Glan-Thompson prism, is the best choice for the primary optical element of the polarimeter to obtain the best possible accuracy in measurements of linear polarization. A linear achromatic quarter-wave plate will be inserted in front of the polarizer for measurements of circular polarization. To avoid modulation of the polarization signals by the optics following the polarimeter, a second achromatic quarter-wave plate will be attached to the polarizer-analyzer so that circularly polarized light will always be transmitted to the following optics. As the polarizer is rotated in fixed 45° steps and the quarter-wave plate is inserted and removed, the degree of elliptical polarization can be measured by analyzing the different transmitted intensities. The Stokes parameters I , Q , and U are measured with the quarter-wave plate removed from the light path; the parameters I , Q , and V are measured with the quarter-wave plate inserted. The sequence of measurements that are derived from the above configuration is given in Table 21.

Table 21. Sequence of measurements with the polarimeter
(a rotating polarizer and quarter-wave plate).

Angle of polarizer (in degrees)	Intensity measurement	
	Wave plate out	Wave plate in
0	$I+Q$	$I+Q$
45	$I+U$	$I+V$
90	$I-Q$	$I-Q$
135	$I-U$	$I-V$
180	$I+Q$	$I+Q$
225	$I+U$	$I+V$
270	$I-Q$	$I-Q$
315	$I-U$	$I-V$

In practice, only four of the eight sequences listed in the table are required to measure the polarization states; the extra measurements will provide data needed to measure residual instrumental polarization at selected times during the mission. The quarter-wave plate assembly includes an optical window to insure proper focus when the quarter-wave plate is removed. In the following sections we present a summary of the analyses we have carried out that have led to this particular design for the polarimeter and we indicate what we expect its operational characteristics will be.

4.2 Basic Options for the Polarimeter

As described briefly in section II.1, measurement of the vector magnetic field is done by analyzing the state of polarization of light emitted by a spectral line sensitive to the Zeeman effect. The line-of-sight component of the field is derived from measurements of circularly polarized light, while the transverse component comes from measurements of linear polarization. Because of the way the Zeeman effect creates the polarized components of the spectral line, the measured circularly polarized intensities are always much stronger than the linear ones, and thus easier to detect. To measure the vector field, the more difficult measurement of linear polarization must be carried out. Thus the history of the measurements of vector magnetic fields has been one of frustration for observers because of two instrumental effects that degrade the measurement of linear polarization: spurious linear polarization introduced by the instrument (usually the optics), and the transformation of circularly polarized light into linearly polarized light (a process called circular cross talk and caused by an imperfect polarimeter). The instrumental polarization of the foreoptics is discussed in section II.3. Here we will address the goal of designing the best possible polarimeter for measuring linear polarization, one that minimizes circular cross talk.

If we preclude the use of electro-optical modulators such as KD*P crystals as being too unreliable for a long-duration, unmanned space mission, we have two basic options for a polarimeter: a rotating wave plate wheel with a fixed polarizer, or a rotating polarizer with a fixed quarter-wave plate. In the following discussion, we will compare these two designs and perform an error analysis for each one. This error analysis will indicate the accuracies required in the orientation of the fast axis of the wave plates, in the retardation of each wave plate, and in the transmission of the analyzer to meet the goal of reducing the instrumental polarization of the polarimeter below 10^{-4} . As a result of these analyses, we will demonstrate that a polarimeter utilizing a rotating polarizer with a fixed quarter-wave plate is the best design for minimizing circular cross talk, and thus is the choice for the SAMEX polarimeter.

To help visualize how errors in measuring linear polarization are introduced by different designs of polarimeters,

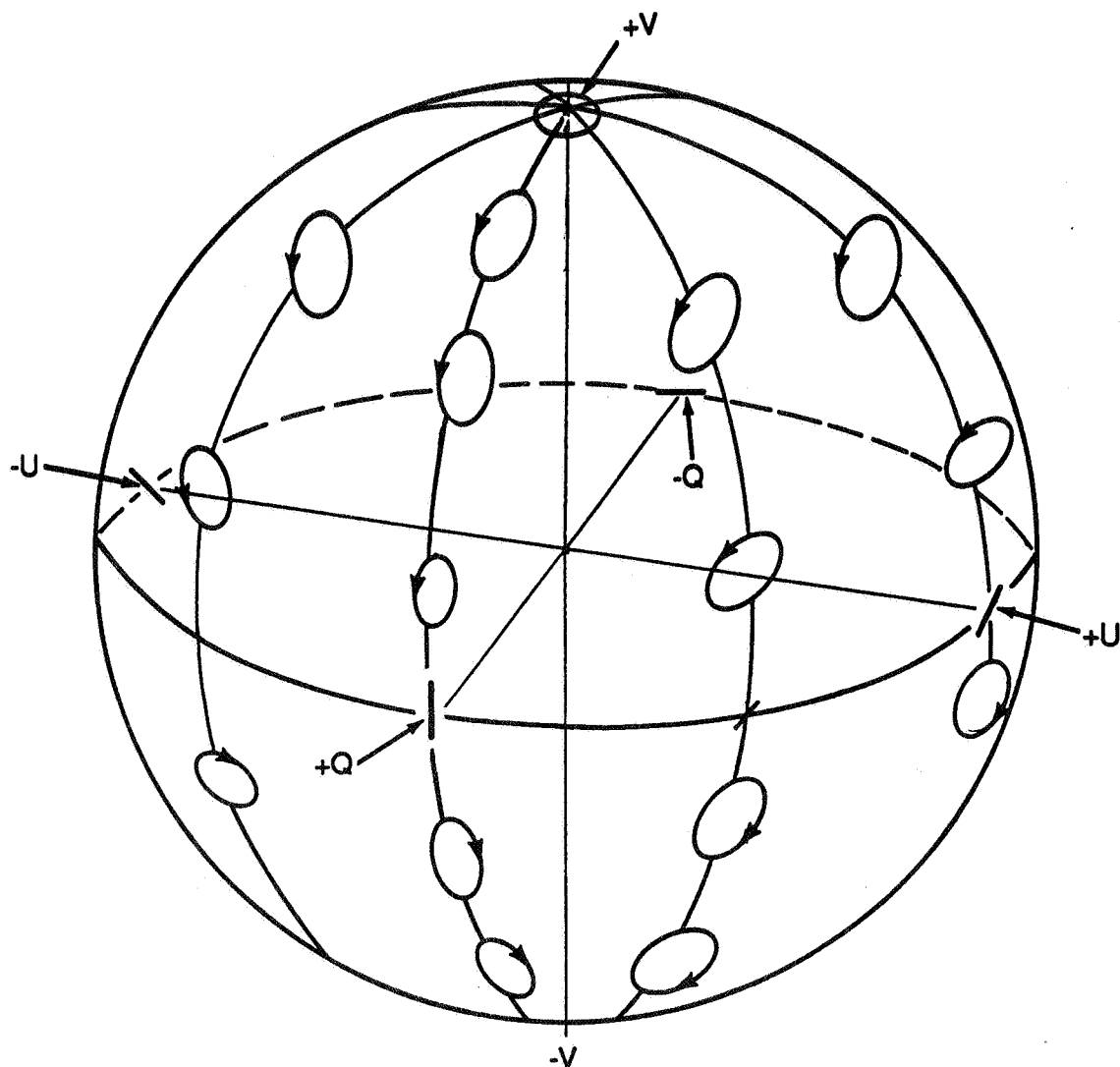
we begin with a short description of the Poincaré sphere and the Stokes vector that provide a visual technique to describe the effects of different retarders on polarized light. (A more detailed discussion of the Poincaré sphere and the Stokes vector can be found in the book Polarized Light, Shurcliff, 1962.)

The Stokes vector $[I, Q, U, V]$ provides a complete description of the state of polarization of a beam of polarized light: the parameter I represents the intensity of the light, Q represents linear polarization oriented at 0° to a defined direction, U represents linear polarization at 45° to that direction, and V represents left circular polarization.

The Poincaré sphere (Figure 32) provides a visual representation for the components of the Stokes vector. Linear polarizations Q and U are mapped along the equator of the Poincaré sphere, and the equatorial positions of linear polarization range from 0° ($+Q$), 90° ($+U$), 180° ($-Q$), and -90° ($-U$). Thus the angular positions on the equator are equal to twice the angular orientations of the linear polarizations (e.g., $+U$ is at the equatorial position of $+90^\circ$ and represents linear polarization at 45°). Left and right circular polarizations ($+V$ and $-V$) are located at the upper and lower poles of the sphere, respectively. The general cases of elliptical polarizations are mapped at all other points on the sphere according to the degree of ellipticity and the orientation of the major axis of the ellipse.

The Poincaré sphere also provides a visual technique to describe the action of a retarder. If the fast axis of a linear retarder is oriented at an angle θ with respect to the $+Q$ orientation, then its effect on any state of polarized light is determined by rotating the Poincaré sphere clockwise about a diameter intersecting the point 2θ on the equator, where the rotation is through an angle given by the retardance of the wave plate. This rotation moves each point on the sphere to a new location which describes its new state of polarization. Thus, transforming left circularly polarized light ($+V$) to linearly polarized light at 0° ($+Q$) requires a clockwise rotation of the sphere through an angle of 90° about a diameter intersecting the point $-U$; this would be done with a retarder with a retardance of 90° (a quarter-wave plate) with a fast axis oriented at -45° to the $+Q$ direction. Similarly, changing left circular polarization to right circular polarization would require a rotation of the sphere through 180° about the same axis, i.e., a half-wave plate with the same orientation of its fast axis.

In measuring one of the components of the Stokes vector with a polarimeter that uses retarding wave plates and an analyzer, the object is to rotate that component to the position on the equator of the Poincaré sphere that represents the orientation of the polarimeter's analyzer. At the same time, the other two components must either remain fixed or be rotated to positions that are located orthogonally with respect to the



POINCARÉ SPHERE

Figure 32. The Poincaré sphere representation of polarized light. Each point on the sphere corresponds to a unique state of polarization. Circular polarizations coincide with the poles of the sphere, and linear polarizations lie on the equator. General, elliptical polarizations lie between the poles and equator; their positions are determined by the ellipticity of the state of polarization and the orientation of the ellipse.

analyzer. Otherwise, the analyzer will also transmit some fraction of the other components, thus creating cross talk in the measurements.

An equivalent representation of the action of a retarder plus analyzer is to rotate the analyzer counterclockwise through an angle equal to the retardance of the wave plate about a diameter determined (as above) by the fast axis of the retarder. Then the position on the Poincaré sphere where the analyzer stops describes the Stokes vector of the light transmitted through both retarder and analyzer. Thus, to measure the +V Stokes component, the analyzer must be rotated up to the north pole of the Poincaré sphere. Any error that results in the analyzer stopping at a position away from the pole introduces other Stokes components; which ones and how much of each are determined by the state of elliptical polarization at the point on the sphere where the analyzer stops.

In the following discussions of the two types of polarimeters, we will use the Poincaré sphere to describe how errors in either the retardation of the wave plate or in the positioning of the fast axis of the plate affect the performance of the polarimeter.

4.3 A Wave Plate Polarimeter

Since polarized light can be completely described using the Stokes vector, only four intensity measurements are needed in theory, one for each component of the vector. However, data reduction can be simplified by over-sampling, and typically six measurements are made so that the polarization of interest can be obtained from ratios of differences to sums:

$$Sq_1 = I + Q$$

$$Q/I = (Sq_1 - Sq_2)/(Sq_1 + Sq_2) , \quad (4.1)$$

$$Sq_2 = I - Q$$

$$Su_1 = I + U$$

$$U/I = (Su_1 - Su_2)/(Su_1 + Su_2) , \quad (4.2)$$

$$Su_2 = I - U$$

$$Sv_1 = I + V$$

$$V/I = (Sv_1 - Sv_2)/(Sv_1 + Sv_2) . \quad (4.3)$$

$$Sv_2 = I - V$$

If a linear polarizer is used as the analyzer and is oriented at 0° on the Poincaré sphere (+Q), then one design that could be used to obtain the six measurements above is the set of six wave plates defined in Figure 33. All six could be mounted on a single "wave plate wheel" that would rotate each plate into the optical light path in succession. Linear measurements (Q/I and U/I) are made by using the four half-wave plates (180° retardation) with their fast axes positioned at 0° , 45° , and $\pm 22.5^\circ$ with respect to the analyzer. (Note that the half-wave plate at 0° has no optical effect on the +Q polarization; a clear glass window could be used for that measurement.) These positions of the half-wave plates result in counterclockwise rotations of the analyzer to the following positions, respectively: +Q (no rotation), -Q, +U, and -U. For measurements of circular polarization, the two other retarders, quarter-wave plates positioned at $\pm 45^\circ$, rotate the analyzer to the north (+V) and south (-V) poles, respectively.

The errors associated with this type of polarimeter are the following (see Figure 34 for representations of these errors on the Poincaré sphere):

1. Misalignment of the fast axis. In the example in Figure 34, a fast axis error rotates the analyzer to a position on the equator slightly displaced from the -U position. This produces linear cross talk in the measurement of the -U parameter by introducing some of the +Q component into the transmitted intensity. There is no circular cross talk because the analyzer remains on the equator.

2. Error in the amount of retardation. For small retardation errors in the linear measurement of -U, the analyzer is rotated just above or below the -U position; the measured intensity is thus elliptically polarized with the axis of the ellipse (almost) parallel to U. Thus the major effect would be circular cross talk (+ or -V, depending on whether the analyzer stops above or below the equator).

3. Elliptical retarder. An elliptical retarder will occur when crystals such as quartz are not cut parallel to the optic axis. The effect of this error is similar to an error in retardation.

To establish a quantitative measure of these errors, we can use the Mueller calculus to describe the effect of different retarding elements on a beam of polarized light (Shurcliff, 1962). In the Mueller calculus, matrix algebra is used to calculate changes in the state of polarization of a beam of light as it passes through a set of optical elements. Each optical element can be mathematically represented by a unique 4 by 4 matrix; the entire set of optical elements can then be represented by a single matrix derived from the product of the individual matrices. The final Stokes vector is obtained by matrix multiplication of the matrix describing the optical system

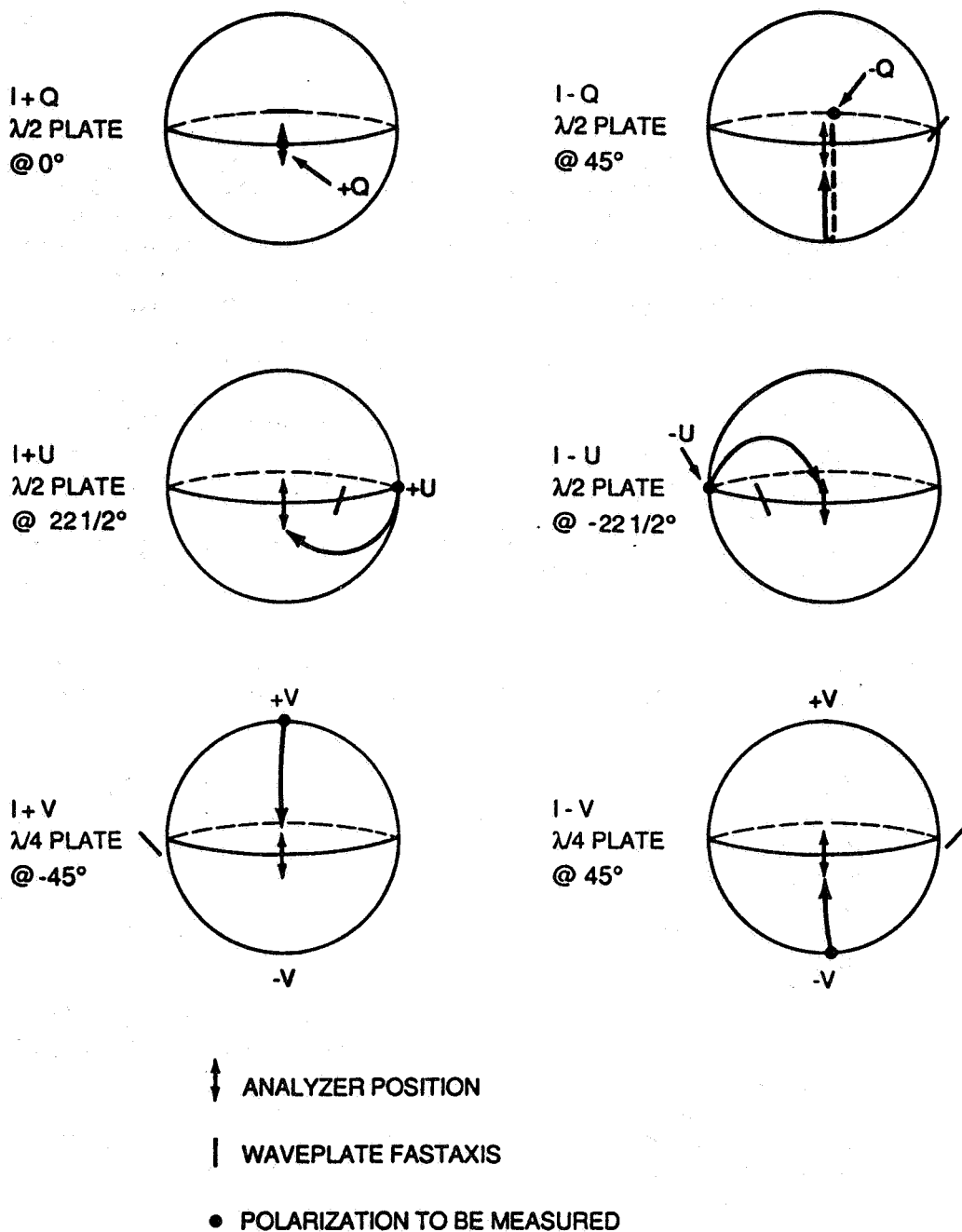


Figure 33. Poincaré sphere representation of the measurements of the Stokes parameters using six wave plates. Each measurement with a wave plate is equivalent to a rotation of the sphere clockwise through an angle given by the retardance of the wave plate about a diameter determined by the orientation of the plate's fast axis. The goal of the wave plate is to move (i.e., transform) each unique polarization state as defined by a Stokes parameter to the position (or state) of the analyzer.

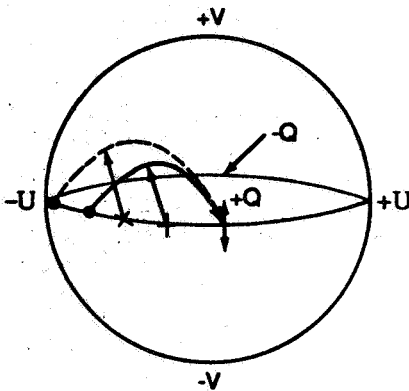
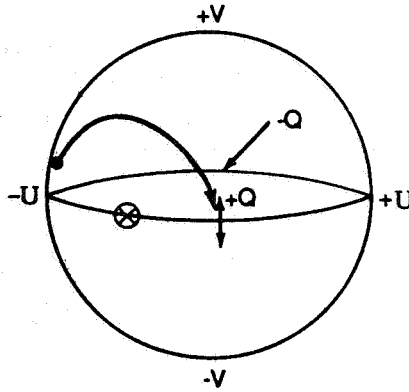
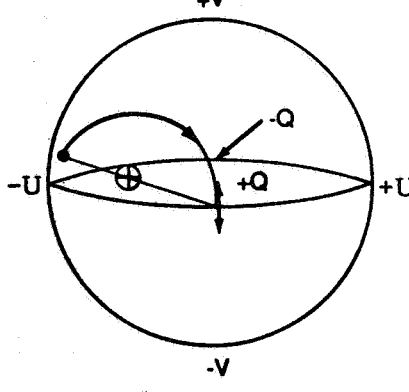
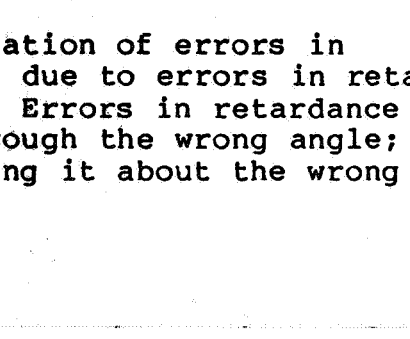
SYMBOL	DESCRIPTION	STOKES VECTOR	POINCARÉ SPHERE
x	HALF-WAVE PLATE WITH NO ERROR	$\begin{bmatrix} I \\ 0 \\ -U \\ 0 \end{bmatrix}$	
+	HALF-WAVE PLATE WITH ERROR IN FAST AXIS	$\begin{bmatrix} I \\ +C_1 Q \\ -C_2 U \\ 0 \end{bmatrix}$	
⊗	HALF-WAVE PLATE WITH SMALL ERROR IN RETARDATION	$\begin{bmatrix} I \\ 0 \\ -C_2 U \\ +C_3 V \end{bmatrix}$	
⊕	ELLIPTICAL RETARDER WITH HALF-WAVE RETARDANCE	$\begin{bmatrix} I \\ 0 \\ -C_2 U \\ +C_3 V \end{bmatrix}$	

Figure 34. Poincaré sphere representation of errors in measurements of the Stokes parameters due to errors in retardances and alignments of wave plates. Errors in retardance are equivalent to rotating the sphere through the wrong angle; errors in alignment are equivalent to rotating it about the wrong axis.

with the Stokes vector describing the initial state of polarization of the incident light. Errors in the optical elements can be represented by introducing error terms in the corresponding Mueller matrix.

The following format will be used to describe the optical performance of a particular polarization measurement:

$$S_x = A_0 I + A_1 Q + A_2 U + A_3 V ,$$

where S_x is a particular polarization measurement (e.g., S_{q1}) and A_0 , A_1 , A_2 , and A_3 are the coefficients that describe the actual performance of the polarimeter. They are the terms in the top row of the resulting Mueller matrix that describes the net effect of all optical elements used in a particular polarization measurement. If the optical elements are perfect retarders whose optical axes are perfectly aligned, the polarization measurements S_x will reduce to those given in Equations (1), (2), and (3). On the other hand, if the optical elements are imperfect, additional Stokes components will be mixed into each of the expressions for S_x , i.e., cross talk occurs, and errors arise in the measurement of the incident Stokes vector.

In presenting results of our analyses in tabular form, we will use normalized coefficients C_i , where

$$C_1 = A_1/A_0,$$

$$C_2 = A_2/A_0,$$

and

$$C_3 = A_3/A_0.$$

In Table 22 we show how a 0.10° error in retardance affects the Stokes coefficients C_i in each of the six measurements in Figure 33 that are made to derive the Stokes components. The retardance error of 0.10° is taken to be a positive error (e.g., the retardance of a half-wave plate is 180° plus the error term). If the error in retardance is negative, the signs of the coefficients change. The sign of a coefficient is important since it can have two different effects on the Stokes components. For example, suppose the light from the Sun incident on the polarimeter is partially circularly polarized and partially linearly polarized with no Q component, and suppose the polarimeter is operated in the mode to measure + and -U.

Table 22. Stokes coefficients for a 0.10° error in wave plate retardation for a rotating wave plate polarimeter.

Polarization Measurement	Coefficients of Stokes Components		
	C_1	C_2	C_3
Sq_1	1.00000	0.000000	0.000000
Sq_2	-0.999998	0.000000	0.001745
Su_1	0.000000	0.999998	0.001234
Su_2	0.000000	-0.999998	-0.001234
Sv_1	0.001745	0.000000	0.999998
Sv_2	0.001745	0.000000	-0.999998

First, let the two half-wave plates at $+$ and -22.5° each have an error in retardance of $+0.10^\circ$. Then in the measurements of $+U$, the analyzer will be rotated just above $+U$ on the Poincare sphere, and just below $-U$ in the measurement of $-U$. Therefore the polarization measured for the U component would be:

$$\begin{aligned}
 U'/I' &= (Su_1 - Su_2)/(Su_1 + Su_2) \\
 &= [(I + C_2U + C_3V) - (I - C_2U - C_3V)] / [(I + C_2U + C_3V) + (I - C_2U - C_3V)], \\
 &= (C_2 U + C_3 V)/I.
 \end{aligned}$$

This is an example of circular cross talk in the linear measurement, and in this example, it increases the measured magnitude of the U Stokes component. On the other hand, if the error in the $+22.5^\circ$ half-wave plate is positive but negative in the other plate, the polarization measured would be:

$$\begin{aligned}
 U'/I' &= \\
 &= [(I + C_2U + C_3V) - (I - C_2U + C_3V)] / [(I + C_2U + C_3V) + (I - C_2U + C_3V)] \\
 &= (C_2U)/(I + C_3V).
 \end{aligned}$$

In this case the magnitude of the linear polarization is reduced by the circular cross talk.

With a design goal of 1×10^{-4} for the polarization accuracy, the data in Tables 22 and 23 indicate that a 0.10° error in the retardance (Table 22) or in the fast axis (Table 23) is too large. This magnitude of error in retardance would lead to cross talk on the order of 2×10^{-3} ; the same error in the alignment of the fast axis produces cross talk of 7×10^{-3} . Both of these would preclude a polarization sensitivity of 10^{-4} .

Table 23. Stokes coefficients for a 0.10° error in the fast axis alignment of wave plates for a rotating wave plate polarimeter.

Polarization Measurement	Coefficients of Stokes Components		
	C_1	C_2	C_3
Sq_1	0.999875	0.006981	0.000000
Sq_2	-0.999975	-0.006981	0.000000
Su_1	-0.006981	0.999975	0.000000
Su_2	0.006981	0.999975	0.000000
Sv_1	0.000012	0.003490	0.999993
Sv_2	-0.000012	-0.003490	-0.999993

These specifications on the retardance and alignment accuracies are tempered somewhat by the fact that the maximum polarization from an active region on the Sun is not 100% but is generally about 35% for circularly polarized light and about 25% for linear polarization. Using 35% as the maximum polarization expected, the coefficient C_3 must be less than 3×10^{-4} to eliminate circular cross talk, and C_1 and C_2 must be less than 4×10^{-4} to eliminate any linear cross talk. These specifications on the coefficients still require a retardation error less than 0.025° and a fast axis error less than 0.01° . Unfortunately these requirements on the wave plates are much too stringent to be physically attainable with today's technology. Experience gained at MSFC in aligning wave plates indicates an uncertainty in the alignment of the fast axis of $\pm 0.3^\circ$ can be expected. Others (e.g., Meadowlark Optics) quote an uncertainty of $\pm 0.18^\circ$ which is still a factor of 8 times the accuracy required to meet the design goal of less than 10^{-4} in instrumental polarization.

With the polarimeter just described, each measurement for a Stokes component is done with a different set of wave plates so that each measurement has different fast-axis errors, retardation errors, errors associated with the internal properties of the birefringent material (Table 24), and errors in the optical manufacturing of the wave plate (elliptical retarder). A better approach than the "wave plate wheel" is to use a single half-wave plate for the linear measurements and rotate it to four different positions, instead of using four different, fixed half-wave plates. Similarly, a single rotating quarter-wave plate is used for the circular measurements. In this design, the number of possible sources of errors is reduced since fewer optical elements are used in the polarization measurements. However, there will still be retardation errors for the two wave plates to consider along with the problem of alignments of the fast axes in each of the six positions of the rotating plates. The magnitudes of these errors will be similar to those given in Tables 22 and 23.

Another source of error that will degrade the performance of this type of polarimeter is that associated with the analyzer. These errors will be discussed in the next section.

Since the design goal of a polarization sensitivity of 1×10^{-4} cannot be obtained with either fixed or rotating wave plates, a different approach must be taken. In the next section, the concept of a polarimeter using a rotating polarizer is discussed, and the polarization errors associated with this approach are analyzed.

Table 24. Properties of wave plate materials.

-
- A. Plastic film (e.g., cellulose nitrate)
1. Inexpensive to manufacture
 2. Achromatic over several hundred angstroms
 3. Each sheet must be evaluated individually since the birefringence depends on the manufacturing process.
 4. Light level must not be higher than 50 watts/cm².
 5. Temperature must be lower than 140° F (Ref. 4).
 6. While the magnitude of the retardation is uniform over a large sheet, the orientation of the fast axis may vary as much as 1.5° (Ref. 3).
- B. Apophyllite
1. Birefringence is directly proportional to wavelength.
 2. Its small birefringence allows construction of a zero-order wave plate from a single piece.
 3. Used to manufacture achromatic wave plates
- C. Quartz (zero-order wave plates)
1. At 5250 Å there is a 0.014° change in retardance per angstrom (Ref. 1). Over the wavelength range of the SAMEX magnetograph, (5244-5254 Å), this would represent an error of $\pm 0.060^\circ$ in retardation.
 2. Zero-order temperature dependence: 0.0091° per C° (Ref. 2).
- D. Compound Plate
1. Materials consist of a positive (quartz) and a negative (calcite) birefringent crystal to make an achromatic wave plate.
 2. Additional materials may be used to correct for temperature and wavelength variations.

References:

1. Optics Handbook (pp. 10-117; figure 63)
 2. Optics Handbook (pp. 10-112)
 3. Optics Handbook (pp. 10-118)
 4. Meadowlark Optics sales literature
-

4.4 A Rotating Polarizer as a Polarimeter

As we have indicated previously, in observations of solar magnetic fields the circularly polarized intensities produced by the Zeeman effect are always much stronger than the linear ones. Because the linear signals are so much weaker, the polarimeter should be designed to minimize circular cross talk into the linear signal. With reference to the Poincare sphere, the best way to eliminate circular cross talk is to avoid optical elements that rotate circular polarizations toward the equator.

This can be done by using a rotating polarizer and making measurements at angles of 0° , 45° , 90° , and 135° . On the Poincare sphere, this is equivalent to putting an analyzer on the equator at $+Q$, $+U$, $-Q$, and $-U$. To measure the circular Stokes component, a quarter-wave plate with its fast axis at 0° must be inserted into the optical train. This plate will rotate the $+$ ($-$) V Stokes component to the $+$ ($-$) U position; measurements are then made with the rotating polarizer at angles of 45° and 135° . The Poincare sphere representation in Figure 35 demonstrates how such a rotating analyzer is used to measure the Stokes vector. As can be seen in Figure 35, the elimination of wave plates with retardation errors excludes any possibility of introducing circular cross talk into the linear measurements. Moreover, any error associated with the alignment of the transmission axis of the analyzer has the same effect as an error in positioning the fast axis of a half-wave plate: it will introduce linear cross talk into the linear measurement, and this is a much weaker source of instrumental polarization compared to circular cross talk.

In considering the problems associated with optical alignments, it is technically quite easy to align a polarizer to a position of minimum transmission (extinction) and then mechanically rotate it to the proper positions (e.g., 45° for the measurement of S_{u1}). On the other hand, it is much more difficult to find the fast axis of a wave plate and then reposition it to its correct orientation. Since the alignment of the fast axis of wave plates depends on the alignment of polarizers, any alignment error in the polarizer will simply propagate into errors in aligning the wave plates. Furthermore, the location of the fast axis of wave plates can be complicated by optical properties of the birefringent material (see Table 24).

In this polarimeter design, the optimum position for the fast axis of the quarter-wave plate used to measure the V Stokes components is at 0° . With this orientation, alignment errors can be minimized since the extinction between the analyzer and a calibration polarizer can be used for the alignment and no subsequent repositioning of the quarter-wave plate is necessary which could produce a mechanical alignment error.

Since the polarimeter can be rotated through 360° , measurements can be made at several additional positions of the rotating polarizer and used for calibrations and redundancy. The polarization measurements that will be used in our subsequent discussions are defined with respect to the angular position of the polarizer (given in degrees) in the following manner:

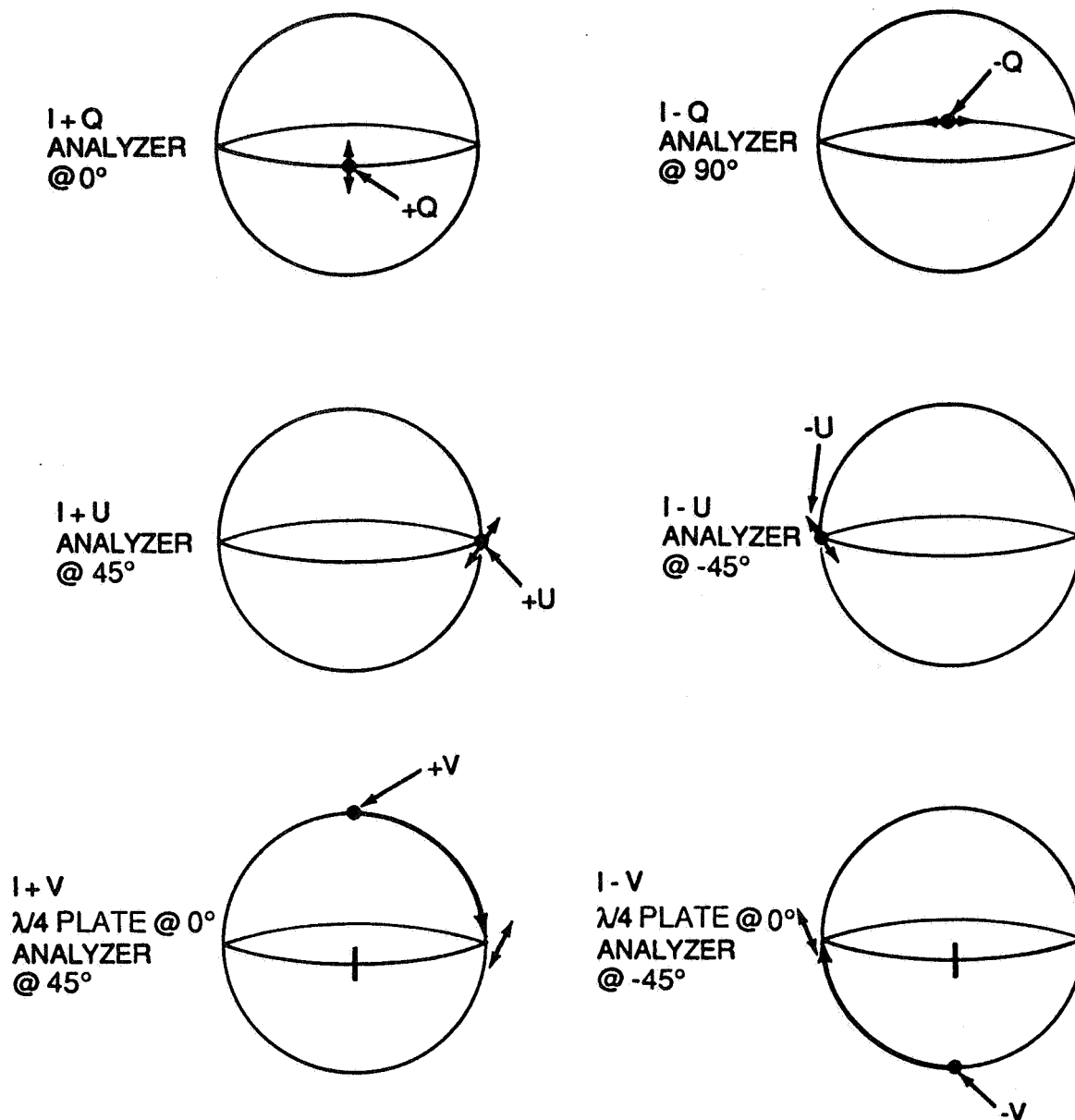


Figure 35. Poincaré sphere representation of the measurements of the Stokes parameters using a rotating polarizer. A quarter-wave plate is needed to measure the circular (V) parameters with the polarizer.

1. Clear window in front of the analyzer

Polarization measurement	Stokes component	Polarizer angle (degrees)
Sq ₁	I+Q	0
Sq ₂	I-Q	90
Sq ₃	I+Q	180
Sq ₄	I-Q	270
Su ₁	I+U	45
Su ₂	I-U	135
Su ₃	I+U	225
Su ₄	I-U	315

2. Quarter-wave plate in front of the analyzer (with fast axis at 0° on the equator of the Poincaré sphere)

Polarization measurement	Stokes component	Polarizer angle (degrees)
Sq ₅	I+Q	0
Sq ₆	I-Q	90
Sq ₇	I+Q	180
Sq ₈	I-Q	270
Sv ₁	I+V	45
Sv ₂	I-V	135
Sv ₃	I+V	225
Sv ₄	I-V	315

This particular observational scheme will produce a set of 16 measurements that correspond to the 16 positions of the polarizer. However, from the point of view of measuring the Stokes components, only 8 of the measurements are unique. The redundant measurements are:

$$\begin{aligned}
 Sq_1 &= Sq_3 \\
 Sq_2 &= Sq_4 \\
 Sq_5 &= Sq_7 \\
 Sq_6 &= Sq_8 \\
 Su_1 &= Su_3 \\
 Su_2 &= Su_4 \\
 Sv_1 &= Sv_3 \\
 Sv_2 &= Sv_4
 \end{aligned}$$

The extra measurements can be used to detect and correct for image displacements created by the rotating polarimeter. This will be discussed in the section on calibration.

There still are errors associated with this polarimeter design because polarizers are not perfect. Since the polarization sensitivity that this design can have is limited by the optical performance of the polarizer, we must address the

errors associated with imperfect dichroic sheet polarizers and birefringent polarizers.

The errors associated with the analyzer, described in terms of its principal transmittances (defined below), are difficult to visualize with the Poincare sphere. Therefore numerical simulations using Mueller matrices and the Stokes vector will be used to demonstrate these sources of error.

The principal transmittances of a polarizer are defined to be k_1 , the ratio of the transmitted light to the incident linearly polarized light when the orientation of the polarizer is set to maximize the transmission, and k_2 , the ratio of the transmitted light to the incident linearly polarized light when the orientation of the polarizer is set to minimize the transmission. For an ideal polarizer, $k_1 = 1$ and $k_2 = 0$.

Contrast ratio is another term that is used to describe the optical performance of polarizers. For two identical polarizers the contrast ratio is defined as the ratio of the intensity of light transmitted through parallel polarizers to the intensity of light transmitted through crossed polarizers; in terms of transmittances k_1 and k_2 , the contrast ratio (CR) is given by the expression

$$CR = (0.5) (k_1^2 + k_2^2) / (k_1 k_2) .$$

In Table 25 we list values of transmittances and contrast ratios for some standard polarizers.

Table 25. Transmittance and contrast ratio for different polarizers (all values are for a wavelength of 5250 Å).

Polarizer	Transmittance		Contrast Ratio
	k_1	k_2	
Polaroid HN22	0.53456	1.99526E-5	13403 : 1
Polaroid HN32	0.69662	5.24807E-5	6636 : 1
Polaroid HN38	0.83176	1.65959E-4	729 : 1
Glan-Thompson	0.95	1.05263E-6	475000 : 1

In terms of the transmittances k_1 and k_2 , the expression for the polarization measurements S_x is easily derived for a non-ideal rotating polarizer with its axis at an angle θ :

$$S_x = I (k_1 + k_2) + Q (k_1 - k_2) \cos 2\theta + U (k_1 - k_2) \sin 2\theta .$$

Thus the linear polarization measurements now become:

<u>Polarization measurement</u>	<u>Polarizer position (degrees)</u>	<u>Stokes component</u>
Sq ₁	0	$(k_1 + k_2)I + (k_1 - k_2)Q$
Sq ₂	90	$(k_1 + k_2)I - (k_1 - k_2)Q$
Su ₁	45	$(k_1 + k_2)I + (k_1 - k_2)U$
Su ₂	135	$(k_1 + k_2)I - (k_1 - k_2)U$

In Table 26, the Stokes coefficients are given for a non-ideal Glan Thompson polarizer. From these data, it is quite evident that there can be no circular cross talk with this type polarimeter, whether or not the polarizer is an ideal one. However, as we show in section 4.5, if there are polarizing optics following this polarimeter, the situation changes dramatically!

Table 26. Stokes coefficients for a Glan-Thompson polarizer followed by non-polarizing optics.

Measurement	<u>Coefficients of Stokes Components</u>		
	<u>C₁</u>	<u>C₂</u>	<u>C₃</u>
Sq ₁	0.999998	0.000000	0.000000
Sq ₂	-0.999998	0.000000	0.000000
Su ₁	0.000000	0.999998	0.000000
Su ₂	0.000000	-0.999998	0.000000

The fixed quarter-wave plate used in the measurements of circular polarizations (Sv) will be another source of error. The errors associated with this wave plate cannot corrupt the primary linear measurements Sq_{1,2} and Su_{1,2} since the plate is not in the optical light path during their measurements. However, both retardation and alignment errors in the quarter-wave plate will produce linear crosstalk (C₁ and C₂, Tables 27 and 28) in the circular measurements (Sv_{1,2}). Note that a retardation error will not affect the secondary linear measurements Sq_{5,6}, either,

if the quarter-wave plate's fast axis is aligned exactly at 0° - the position of the analyzer will still be orthogonal to any circular polarization. To meet the design goal of a polarimetric sensitivity of 1×10^{-4} , the retardation error of the wave plate must be less than 0.03° and its alignment error less than 0.01° . Again, the experience of researchers at MSFC indicates that the alignment and retardation errors of a quarter-wave plate can at best be reduced to $\pm 0.1^\circ$. Therefore the required accuracies represent an order of magnitude better than those presently obtained. This means that some data analysis techniques will have to be developed to correct for these errors; these analysis techniques will be discussed in the section on calibration. It should be remembered, however, that these errors are second-order effects of linear cross talk (C_1, C_2) into the circular measurements (S_v).

Table 27. Stokes coefficients as a function of retardation errors in the quarter-wave plate used in the circular measurements (S_{v1}, S_{v2}).

Measurement	Retardation Error (degree)	Coefficients of Stokes Components		
		C_1	C_2	C_3
S_{q5}	0.100	1.000000	0.000000	0.000000
S_{q6}	0.100	-1.000000	0.000000	0.000000
S_{v1}	0.100	0.000000	-0.001745	0.999998
S_{v2}	0.100	0.000000	0.001745	-0.999998
S_{v1}	0.01	0.000000	-0.000174	0.999999
S_{v2}	0.01	0.000000	0.000174	-0.999999

Table 28. Stokes coefficients as a function of alignment errors in the quarter-wave plate used in the circular measurements (Sv_1 , Sv_2).

Measurement	Fast Axis Error (degree)	Coefficients of Stokes Components		
		C_1	C_2	C_3
Sq_5	0.100	0.999988	0.003490	-0.003490
Sq_6	0.100	-0.999988	-0.003490	0.003490
Sv_1	0.100	0.003490	0.000013	0.999993
Sv_2	0.100	-0.003490	-0.000013	-0.999993
Sq_5	0.01	0.999999	0.000349	-0.000349
Sq_6	0.01	-0.999999	-0.000349	0.000349
Sv_1	0.01	0.000349	0.000000	0.999999
Sv_2	0.01	-0.000349	-0.000000	-0.999999

4.5 Interface Requirements for a Rotating Polarizer

With a rotating polarizer as the polarimeter of choice, the light transmitted through the rest of the optics will be linearly polarized and the plane of this linear polarization will rotate. This rotation may introduce further errors in the polarimetric analysis, depending on the optical elements following the polarimeter, especially the spectral filter and any folding optics required. To address this subject, we will discuss the problems associated with interfacing this polarimeter design with the spectral filter and other optical elements that follow the polarimeter, the instrumental errors introduced by that interface, and the calibrations required to correct for these errors.

In the ideal situation, the following optics would not be sensitive to the change in the azimuth of the linear polarization. But, since this instrument will probably have folding mirrors to reduce its length, some sensitivity to changing linear polarization is expected. In addition, if the spectral filter is a birefringent type, its first element is a polarizer, so each change in the azimuth of the linear polarization from the polarimeter would produce a change in the transmission of the filter. Moreover, solid-etalon, Fabry-Perot filters may exhibit birefringence and thus be sensitive to changes in the incident linear polarization.

To minimize these problems, a second quarter-wave plate that has its fast axis aligned at 45° to the analyzer and that rotates with the analyzer will be employed. This arrangement will change the linear polarization from the polarizer into circular polarization (Figure 36). The design of the SAMEX polarimeter thus consists of two rotating elements, a linear polarizer and quarter-wave retarder attached to the polarizer. We must now investigate the errors associated with this design as it interfaces with the following optics. The requirements on the rotating quarter-wave plate will depend on the polarization sensitivity of the following optics; the worst case would be a linear polarizer at the entrance of the spectral filter. We will treat each element separately to isolate the errors associated with the polarizer, retarder and following optics (polarizing or non-polarizing).

(1) Case of a non-ideal rotating polarizer and perfect (rotating) quarter-wave plate followed by a filter with an ideal polarizer as its first element:

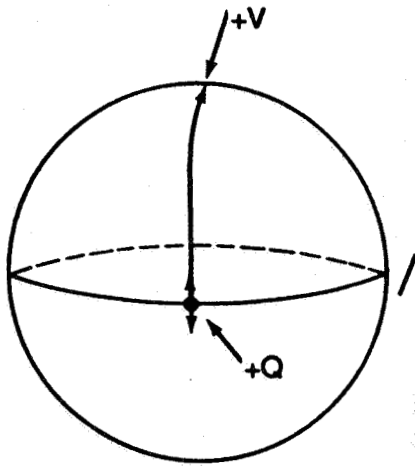
The effect of a non-ideal polarizer (i.e., $k_2 \neq 0$) coupled with ideal polarizing elements in the following optics (i.e., quarter-wave plate and birefringent filter) is to produce some circular and linear crosstalk. Specifically, in the measurements of the Q and U Stokes parameters with the rotating polarizer at 0° and 45° , respectively, the polarization measurements S_x are

$$S_{q1} = I (k_1 + k_2) + Q (k_1 - k_2) - 2V (k_1 k_2)^{1/2}$$

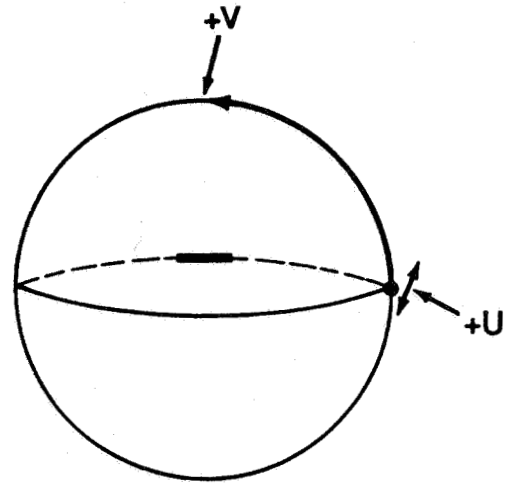
and

$$S_{u1} = I (k_1 + k_2) + 2Q (k_1 k_2)^{1/2} + U (k_1 - k_2).$$

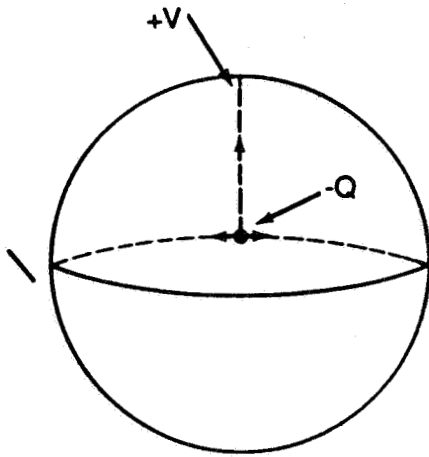
These equations show that there will be circular cross talk in the measurement of the Q component. In Tables 29 through 32 we give the coefficients of the Stokes components for various non-ideal polarizers. To produce circular crosstalk that is less than the design goal for the polarimetric sensitivity of the polarimeter (10^{-4}) requires the circular coefficients (C_3) to be less than 0.0003 for the S_q and S_u measurements. Looking at the data in Tables 29 through 32, we find that the smallest coefficient is 0.002 (for a Glan-Thompson polarizer, Table 32), which is a factor of 10 larger than the design goal. Although the Glan Thompson has a high contrast ratio, the leakage of circular polarization through the analyzer is proportional to the square root of the product of the transmittances. To meet the design goals, a Glan-Thompson polarizer with a minor transmittance (k_2) of 5×10^{-8} would be required. It must be noted that in these calculations we have assumed that the spectral filter following the polarimeter in the optics is a perfect polarizer and thus causes no further degradation in the polarimetric analysis; we will discuss the problem of interfacing the polarimeter with a non-ideal polarizing filter later in a subsequent section.



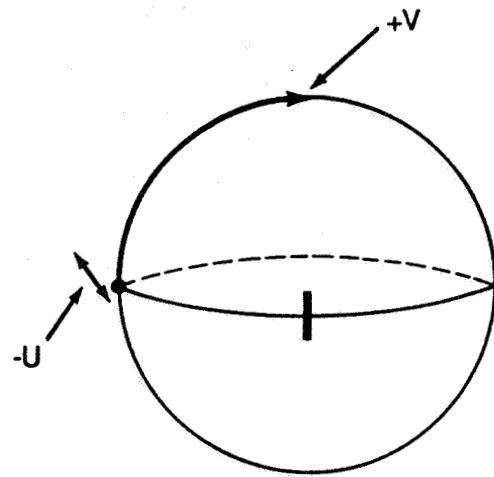
ANALYZER AT 0° , $\lambda/4$ PLATE AT 45°



ANALYZER AT 45° , $\lambda/4$ PLATE AT 90°



ANALYZER AT 90° , $\lambda/4$ PLATE AT 135°



ANALYZER AT 135° , $\lambda/4$ PLATE AT 0° (180°)

Figure 36. Poincaré sphere representation of the effect of a quarter-wave plate rotating with the analyzer. The quarter-wave plate is added to the polarimeter described in Figure 35 to minimize the effects caused by linear polarizers or birefringent elements in the optics following the polarimeter. The effect of the wave plate is to transform all light coming through the polarimeter into circularly polarized light.

Table 29. Stokes coefficients for a single HN22 polarizer followed by ideal rotating quarter-wave plate and filter polarizer (maximum transmission through polarimeter and filter polarizer = 13%).

Measurement	Coefficients of Stokes Components		
	C_1	C_2	C_3
Sq ₁	0.999925	0.000000	-0.012218
Sq ₂	-0.999925	0.000000	0.012218
Sq ₅	0.999925	0.012218	0.000000
Sq ₆	-0.999925	-0.012218	0.000000
Su ₁	0.012218	0.999925	0.000000
Su ₂	0.012218	-0.999925	0.000000
Sv ₁	0.012218	0.000000	0.999925
Sv ₂	0.012218	0.000000	-0.999925

Table 30. Stokes coefficients for a single HN32 polarizer followed by ideal rotating quarter-wave plate and filter polarizer (maximum transmission through polarimeter and filter polarizer = 17%).

Measurement	Coefficients of Stokes Components		
	C_1	C_2	C_3
Sq ₁	0.999849	0.000000	-0.017358
Sq ₂	-0.999849	0.000000	0.017358
Sq ₅	0.999849	0.017358	0.000000
Sq ₆	-0.999849	-0.017358	0.000000
Su ₁	0.017358	0.999849	0.000000
Su ₂	0.017358	-0.999849	0.000000
Sv ₁	0.017358	0.000000	0.999849
Sv ₂	0.017358	0.000000	-0.999849

Table 31. Stokes coefficients for a single HN38 polarizer followed by ideal rotating quarter-wave plate and filter polarizer (maximum transmission through polarimeter and filter polarizer = 20%).

Measurement	Coefficients of Stokes Components		
	C_1	C_2	C_3
Sq ₁	0.998679	0.000000	-0.051374
Sq ₂	-0.998679	0.000000	0.051374
Sq ₅	0.998679	0.051374	0.000000
Sq ₆	-0.998679	-0.051374	0.000000
Su ₁	0.051374	0.998679	0.000000
Su ₂	0.051374	-0.998679	0.000000
Sv ₁	0.051374	0.000000	0.998679
Sv ₂	0.051374	0.000000	-0.998679

Table 32. Stokes coefficients for a Glan-Thompson polarizer followed by ideal rotating quarter-wave plate and filter polarizer (maximum transmission through polarimeter and filter polarizer = 24%).

Measurement	Coefficients of Stokes Components		
	C_1	C_2	C_3
Sq ₁	0.999997	0.000000	-0.002052
Sq ₂	-0.999997	0.000000	0.002052
Sq ₅	0.999997	0.002052	0.000000
Sq ₆	-0.999997	-0.002052	0.000000
Su ₁	0.002052	0.999997	0.000000
Su ₂	0.002052	-0.999997	0.000000
Sv ₁	0.002052	0.000000	0.999997
Sv ₂	0.002052	0.000000	-0.999997

Another approach using existing technology is to use two parallel polarizers. Then, the circular coefficient (C_3) would be proportional to the product of the transmittances. This approach is demonstrated by the data in Tables 33 and 34. The disadvantage with this approach is the decrease in transmission through the polarimeter. Thus, two HN32 Polaroids would meet the design goal with a circular coefficient (C_3) equal to 0.00015, but the transmission would be reduced from 17% to 12%. To increase the transmission, a Glan-Thompson and a HN38 Polaroid combination could be used; this is demonstrated in the data in Table 35. The circular coefficient (C_3) for this combination is 100 times better than that for a single Glan-Thompson polarizer, and the net transmission is 20%, a factor of almost 2 better than the 12% transmission of the two HN32 Polaroids.

Although Glan-Thompson polarizers have a high transmission and a large contrast ratio, there are other problems associated with them. Some of the more common problems are listed in Table 36. In the section on calibrations we will discuss how the errors stemming from some of these problems can be corrected. In Appendix A, parameters important in the fabrication of high-quality Glan-Thompson polarizers are discussed.

Table 33. Stokes coefficients for two HN22 polarizers followed by ideal rotating quarter-wave plate and filter polarizer (maximum transmission through polarimeter and filter polarizer = 7%).

Measurement	Coefficients of Stokes Components		
	C_1	C_2	C_3
Sq_1	0.999999	0.000000	-0.000074
Sq_2	-0.999999	0.000000	0.000074
Sq_5	0.999999	0.000074	0.000000
Sq_6	-0.999999	-0.000074	0.000000
Su_1	0.000074	0.999999	0.000000
Su_2	0.000074	-0.999999	0.000000
Sv_1	0.000074	0.000000	0.999999
Sv_2	0.000074	0.000000	-0.999999

Table 34. Stokes coefficients for two HN32 polarizers followed by ideal rotating quarter-wave plate and filter polarizer (maximum transmission through polarimeter and filter polarizer = 12%).

Measurement	Coefficients of Stokes Components		
	C_1	C_2	C_3
Sq ₁	0.999999	0.000000	-0.000150
Sq ₂	-0.999999	0.000000	0.000150
Sq ₅	0.999999	0.000150	0.000000
Sq ₆	-0.999999	-0.000150	0.000000
Su ₁	0.000150	0.999999	0.000000
Su ₂	0.000150	-0.999999	0.000000
Sv ₁	0.000150	0.000000	0.999999
Sv ₂	0.000150	0.000000	-0.999999

Table 35. Stokes coefficients for two parallel polarizers: a Glan-Thompson and a HN38 Polaroid followed by ideal rotating quarter-wave plate and filter polarizer (maximum transmission through polarimeter and filter polarizer = 20%).

Measurement	Coefficients of Stokes Components		
	C_1	C_2	C_3
Sq ₁	0.999999	0.000000	-0.000029
Sq ₂	-0.999999	0.000000	0.000029
Sq ₅	0.999999	0.000029	0.000000
Sq ₆	-0.999999	-0.000029	0.000000
Su ₁	0.000029	0.999999	0.000000
Su ₂	0.000029	-0.999999	0.000000
Sv ₁	0.000029	0.000000	0.999999
Sv ₂	0.000029	0.000000	-0.999999

Table 36. Common defects in Glan-Thompson prism polarizers.

1. Squirm: the displacement of an axial beam as the polarizer is rotated. This occurs when the optic axes in the two halves are not parallel. It will also occur if the faces of the two prisms are not parallel. Deviations of less than 1 arc minute are possible.

2. Variations in transmittance: variations in transmittance can occur between positions of the polarizer that are 180° apart. Explanations for these variations include the transmission of stray light which may be outside the entrance or exit field angles, scattered light from the surfaces that are to absorb the ordinary rays, and strain birefringence in the calcite.

3. Axis wander: the variation of the azimuth of the transmitted polarization over the polarizer aperture. Typical values are 0.01 - 0.02° , whereas the best prisms can have an axis wander of $\pm 0.003^\circ$.

4. Ellipticity: arc tangent of the ratio of the minor and major axis of the emergent elliptical polarization. The best prisms have an ellipticity of 0.005° . The maximum values are 0.01 - 0.04° .

Proper use of Glan-Thompson prisms:

1. The field angle of the prism should not be exceeded.
 2. The prism should not be the limiting aperture of the optical system.
 3. Baffles should be placed on each side of the polarizer so that stray light will not be transmitted.
-

(2) Case of ideal rotating polarizer and a non-ideal rotating quarter-wave plate followed by a filter with an ideal polarizer as its first element:

We now examine the errors introduced by a quarter-wave plate with errors in its retardation and alignment. The data in Tables 37 and 38 show how errors in the rotating quarter-wave plate affect the Stokes vector. In calculating the data for these two tables, we have assumed an ideal rotating analyzer and entrance polarizer (in the filter) so that the errors introduced by the quarter-wave plate can be isolated and studied. Even with large errors in the retardation and alignment of the quarter-wave plate, the Stokes coefficients (C_1 , C_2 , and C_3) are not affected. However, there will be a bias in the polarization measurements due to the modulation in the transmission coefficient (A_0).

Table 37. Transmission and Stokes coefficients for a 5° retardation error in the rotating quarter-wave plate following an ideal rotating analyzer and preceding an ideal polarizer.

Measurement	Transmission Coefficient A_0	Coefficients of Stokes Components		
		C_1	C_2	C_3
Sq ₁	0.228211	1.000000	0.000000	0.000000
Sq ₂	0.271788	-1.000000	0.000000	0.000000
Sq ₅	0.228211	1.000000	0.000000	0.000000
Sq ₆	0.271788	-1.000000	0.000000	0.000000
Su ₁	0.250000	0.000000	1.000000	0.000000
Su ₂	0.249999	0.000000	-1.000000	0.000000
Sv ₁	0.250000	0.000000	0.000000	1.000000
Sv ₂	0.249999	0.000000	0.000000	-1.000000

Table 38. Transmission and Stokes coefficients for a 5° alignment error in the rotating quarter-wave plate following an ideal rotating analyzer and preceding an ideal polarizer.

Measurement	Transmission Coefficient A_0	Coefficients of Stokes Components		
		C_1	C_2	C_3
Sq ₁	0.257538	1.000000	0.000000	0.000000
Sq ₂	0.242462	1.000000	0.000000	0.000000
Sq ₅	0.257538	1.000000	0.000000	0.000000
Sq ₆	0.242462	-1.000000	0.000000	0.000000
Su ₁	0.207247	0.000000	1.000000	0.000000
Su ₂	0.292752	0.000000	-1.000000	0.000000
Sv ₁	0.207247	0.000000	0.000000	1.000000
Sv ₂	0.292752	0.000000	0.000000	-1.000000

(3) Case of a Glan-Thompson rotating polarizer and a non-ideal rotating quarter-wave plate followed by non-polarizing optics:

If there is no polarization sensitivity in the optics following the polarimeter (which includes the rotating wave plate), then the transmission coefficient will be constant even with a non-ideal rotating polarizer and large errors in the quarter-wave plate; this result is demonstrated by the data in Table 39 for the case of polarization-insensitive optics following the polarimeter. Only when there is a sensitivity to polarized light in the following optics do the errors introduced by the rotating quarter-wave plate become critical. In Table 39, we have assumed that the spectral filter is a Fabry-Perot device with no sensitivity to polarized light and that no folding mirrors are used. In this case, large errors in the quarter-wave plate have no effect on the transmission coefficient (A_0).

Table 39. Transmission and Stokes coefficients for 5° retardation and alignment errors in the rotating quarter-wave plate following a rotating Glan-Thompson polarizer and preceding non-polarizing optics.

Measurement	Transmission Coefficient A_0	Coefficients of Stokes Components		
		C_1	C_2	C_3
Sq ₁	0.475000	0.999997	0.000000	0.000000
Sq ₂	0.475000	-0.999997	0.000000	0.000000
Sq ₅	0.475000	0.999997	0.000000	0.000000
Sq ₆	0.475000	-0.999997	0.000000	0.000000
Su ₁	0.475000	0.000000	0.999997	0.000000
Su ₂	0.475000	0.000000	-0.999997	0.000000
Sv ₁	0.475000	0.000000	0.000000	0.999997
Sv ₂	0.475000	0.000000	0.000000	-0.999997

(4) Case of a Glan-Thompson rotating polarizer and a non-ideal rotating quarter-wave plate followed by a filter with a non-ideal polarizer as its first element:

If a birefringent spectral filter is part of the optics following the polarimeter, the following optics become polarization sensitive since the first optical element of the filter is a polarizer. In Table 40 we show Stokes and transmission coefficients calculated for the case of a birefringent filter with a HN38 entrance polarizer. Since the Glan-Thompson polarizer is not perfect (Table 25), the errors associated with the quarter-wave plate not only affect the transmission coefficient (A_0), but also change the Stokes coefficients (compare Tables 32 and 40). Of course, the actual errors in retardation and alignment are expected to be an order of magnitude smaller than those used in this simulation.

Table 40. Transmission and Stokes coefficients for 5° retardation and alignment errors in the rotating quarter-wave plate following a rotating Glan-Thompson polarizer and preceding a HN38 polarizer.

Measurement	Transmission Coefficient A_0	Coefficients of Stokes Components		
		C_1	C_2	C_3
Sq_1	0.186843	0.999997	0.000413	-0.002183
Sq_2	0.208321	-0.999997	0.000370	0.001958
Sq_5	0.186843	0.999997	0.002183	0.000413
Sq_6	0.208321	-0.999997	0.001958	0.000370
Su_1	0.160863	0.002499	0.999997	-0.000447
Su_2	0.234301	0.001716	-0.999998	0.000307
Sv_1	0.160863	0.002499	0.000447	0.999997
Sv_2	0.234301	0.001716	0.000306	-0.999998

These and similar analyses of all the different effects of non-ideal polarizers and retarders have led to the concept of a hybrid polarimeter designed to interface with a non-ideal birefringent filter. Assuming reasonable values for errors in alignment and retardation for the quarter-wave plate and for transmittances of all polarizers, we find that this hybrid

polarimeter can meet the SAMEX specifications for a polarization accuracy of 10^{-4} in the linear measurements ($Sq_{1,2}$ and $Su_{1,2}$). The data supporting this statement are presented in Table 41.

Table 41. Transmission and Stokes coefficients for a hybrid rotating polarimeter followed by a HN38 entrance polarizer in a birefringent filter. The hybrid polarimeter is composed of a Glan-Thompson and a HN38 polarizer and a quarter-wave plate. A removable quarter-wave plate is inserted for measurements of the V Stokes intensities. Errors in both quarter-wave plates (inserted and rotating) are assumed to be 0.5° in both retardations and alignments.

Measurement	Transmission Coefficient A_0	Coefficients of Stokes Components		
		C_1	C_2	C_3
Sq_1	0.165824	0.999999	0.000000	-0.000029
Sq_2	0.162858	-0.999999	0.000000	0.000029
Sq_5	0.165824	0.999697	0.017327	-0.017452
Sq_6	0.162858	-0.999697	-0.017327	0.017452
Su_1	0.161499	0.000030	0.999999	-0.000000
Su_2	0.167182	0.000030	-0.999999	0.000000
Sv_1	0.161499	0.017327	0.009029	0.999809
Sv_2	0.167182	-0.017268	-0.009028	-0.999810

4.6 Calibrations for Residual Errors

To measure and eliminate residual errors introduced by the polarimeter, we will have to carry out several calibration sequences during the operation of the vector magnetograph. We have already identified several of these calibration procedures and these are outlined in the following discussion.

(1) Residual errors from the rotating quarter-wave plate:

Since the errors in the rotating quarter-wave plate affect all components of the Stokes vector, these errors can be measured and eliminated from the polarimetric data by measuring the intensities transmitted by the polarimeter when unpolarized light is incident on it. Unpolarized light can best be realized by observing in a non-magnetic spectral line or in the solar continuum. Once a complete set of polarization measurements is acquired for unpolarized light, the following equation can be used to correct for any biases created by the rotating quarter-wave plate:

$$P_x = (Sx_1/Cn_1 - Sx_2/Cn_2)/(Sx_1/Cn_1 + Sx_2/Cn_2) .$$

In this equation, P_x is the corrected polarization measurement (x represents q , u , or v), Cn_1 is the correction term for the first polarization measurement Sx_1 , and Cn_2 is the correction term for the second polarization measurement Sx_2 . The correction terms Cn_1 and Cn_2 are derived from the following equations:

$$Cn_1 = T_0 - Cn_2 ,$$

$$Cn_2 = (T_0/2) [1 - (Cx_1 - Cx_2)/(Cx_1 + Cx_2)] ,$$

where Cx is the same polarization measurement for unpolarized light (e.g., $Cq_1 = Sq_1 = A_0 I$), T_0 is the transmission through the polarizing optics and is equal to the product of the transmittances of the polarizers in the instrument:

$$T_0 = (0.5) (k_1 + k_2)(k_3 + k_4) = Cx_1 + Cx_2 .$$

In this last expression, k_1 and k_2 are the principal transmittances of the analyzer, and k_3 and k_4 are the principal transmittances of the following optics.

(2) Residual errors from the Glan-Thompson polarizer:

Although Glan-Thompson polarizers give high transmission levels and contrast ratios, there are certain problems associated with them; these are listed in Table 36. One of these problems which is an area of concern for the SAMEX polarimeter is the displacement of the beam as the polarimeter rotates 360° , referred to as squirm (Table 36). This is a particularly worrisome problem with Glan-Thompson polarizers that have a thick birefringent element. The procedure required to correct for this problem is to place a cross hair at the prime focus of the telescope and record its displacement as the polarimeter rotates. If the "cross hair" is produced by a cross slit which transmits light, it can also be used to test the charge transfer efficiency of the CCD camera system (this will be discussed in the section on the detector system).

The redundant polarization measurements (e.g., Sq_1 and Sq_3) can be used to determine and minimize any variations in transmittance between 180° positions. With the addition of the HN38 Polaroid, errors associated with axis wander or ellipticity will be minimized.

(3) Residual errors in alignments and retardations:

To determine the optical errors of the fixed quarter-wave plate used by the polarimeter in the circular measurements, the redundant linear measurements (Sq_5 and Sq_6) will be used. Assuming that the polarimeter consists of two parallel polarizers, the following procedure would be used:

$$Q_1 = (Sq_1 - Sq_2)/(Sq_1 + Sq_2) = Q/I ,$$

$$Q_2 = (Sq_5 - Sq_6)/(Sq_5 + Sq_6) = (C_1 Q + C_2 U + C_3 V)/I ,$$

$$U_1 = (Su_1 - Su_2)/(Su_1 + Su_2) = U/I ,$$

$$V_1 = (Sv_1 - Sv_2)/(Sv_1 + Sv_2) = (C_1' Q + C_2' U + C_3' V)/I .$$

The residual errors in the Q_1 and U_1 measurements are below the 10^{-4} resolution (Table 35) and will be neglected in the analysis of the fixed quarter-wave plate errors. Simple subtraction of Q_1 and Q_2 will determine if optical errors exist in the fixed quarter-wave plate. Since the fixed quarter-wave plate is not used in the Q_1 and U_1 measurements, those measurements are used to determine if linear relationships exist between them and Q_2 . The Q_1 and Q_2 measurements are compared first since they are measuring the "same" polarization. C_1 represents the relationship between the Q_1 and Q_2 measurements determined by the correlation analysis. If the fast axis error is zero (Table 27), then C_1 should be equal to 1 and the difference between the two measurements would be zero. If C_1 is not 1, then

$$Q_3 = Q_2 - C_1 Q_1 = (C_2 U + C_3 V)/I .$$

The analysis would then look for a correlation between Q_3 and U_1 (C_2) which would leave only the circular cross talk component (see footnote):

$$Q_4 = Q_3 - C_2 U_1 = C_3 V/I .$$

C_1 and C_2 could then be used to determine C_3 by solving for the fast axis and retardation errors (Table 42). C_3 would then be used as a calibration check for the C_1' correction in the circular measurement (V_1).

A process similar to that discussed above would be used to correct for linear cross talk in the circular measurement V_1 :

$$V_2 = V_1 - C_1' Q_1 = (C_2' U + C_3' V)/I ,$$

$$V_3 = V_2 - C_2' U_1 = C_3' V/I .$$

C_1' would be determined from the correlation analysis between V_1 and Q_1 and C_2' from the correlation between V_1 and U_1 . Again C_3' would be determined by using C_1' (which should be equal to C_3) and C_2' to solve for the fast axis and retardation errors.

In summary, although the Q_2 measurement is redundant and the errors in the fixed quarter-wave plate can be determined using the correlation analysis on the V_1 measurement only, the C_2 coefficient is more sensitive to fast axis errors in the fixed quarter-wave plate, while the C_2' coefficient is more sensitive to retardation errors. Although both of these errors are systematic, the retardation error is expected to drift more than the fast axis error, since the latter is mechanically set and should not change with time. However, the wave plate retardance, which varies with temperature, must be controlled to within 0.01° (Table 27) in retardation to achieve the 10^{-4} polarization accuracy; the exact temperature relationship will depend upon the birefringent material used for the wave plate. If the retardation cannot be controlled to the required accuracy, the errors must be corrected using the U_1/V_1 correlation technique. The various cross correlations for analyzing the error sources are summarized in Table 42.

Footnote

A possible extension of this technique which would reduce the errors in evaluating C_2 is the comparison of the polarization signals seen in the red (r) and blue (b) wings (± 90 mÅ from line center) of the magnetic line:

$$Q_3 = (Q_{3b} + Q_{3r}) = C_2 (U_{1b} + U_{1r}) + C_3 (V_{1b} + V_{1r}) .$$

Assuming both symmetric filter and magnetic absorption line profiles, we would have $V_{1b} \approx V_{1r}$ so that the expression for Q_3 would reduce to

$$Q_3 = C_2 (U_{1b} + U_{1r}) + C_3 D(f,B) ,$$

where $D(f,B)$ is nearly zero and represents the differences arising from asymmetries in filter (f) and absorption line (B) profiles. Plotting Q_3 versus $(U_{1b} + U_{1r})$ will allow C_2 to be determined.

Table 42. Cross correlations for error analysis.

Correlated Data	Correlation	Error Source
$Q_1 \times V_1$	C_1'	fast axis
$U_1 \times V_1$	C_2'	retardance
$Q_1 \times Q_2$	C_1	fast axis
$U_1 \times Q_2$	C_2	fast axis

These measurements are used to determine if a fast axis error exists and to check the V_1 correlation measurements.

$$C_1 = \cos(4\theta) (\sin(\delta/2))^2 + (\cos(\delta/2))^2$$

$$C_2 = \sin(4\theta) (\sin(\delta/2))^2$$

$$C_1' = -C_3 = -\sin(2\theta) \sin(\delta)$$

$$C_2' = \cos(2\theta) \sin(\delta)$$

$$C_3' = \cos(\delta)$$

where θ is the fast axis of the fixed quarter-wave plate ($\theta \sim 0^\circ$) and δ is the retardance of the fixed quarter-wave plate ($\delta \sim 90^\circ$).

A calibration wheel assembly in front of the fixed quarter-wave plate could also be used to calibrate and monitor the performance of both the polarimeter and CCD detector. The calibration wheel should include (1) a diffuser and three linear polarizers with axes oriented at 0° , 120° , and 240° for calibrations of linear polarization; (2) a diffuser and linear polarizer and quarter-wave plate for calibrations of circular polarization; and (3) neutral density filters (a set of 1ND, 2ND, 3ND, and Opaque).

From this discussion, we see that this design for the SAMEX polarimeter will allow us to correct for any residual errors using in-flight calibrations and the redundant measurements. In fact, it is this redundancy that assures the possibility of corrections after the instrument is operational (and beyond recovery for further calibrations). In contrast, a polarimeter using fixed wave plates on a rotating wheel provides none of this redundancy. Consequently, all sources of errors must be known before the experiment is flown and the appropriate calibrations made. Otherwise, any errors not accounted for that occur in the actual operation in space cannot be eliminated.

through in-flight calibration. The seriousness of this deficiency can be appreciated when one tries to visualize a situation where all sources of possible errors can be anticipated before launch in an instrument this complex.

4.7 Summary

The data given in Table 41 indicate that the design goal of a polarimetric sensitivity of 1×10^{-4} can be achieved in the linear measurements using a hybrid analyzer. The accuracies required for the fixed quarter-wave plate, a maximum error of 0.02° in retardance and 0.01° in alignment of the fast axis, are not technically feasible at the present time. Thus the errors introduced by these inaccuracies will require in-flight calibrations and corrections to the data through analysis of these calibrations. The redundant linear measurements $Sq_{5,6}$ can be used to estimate, correct and monitor errors associated with the fixed quarter-wave plate.

To minimize instrumental effects, the spectral filter should not be sensitive to polarization. The more sensitive a filter is to polarization, the tighter will be the requirements on the rotating quarter-wave plate following the analyzer. To eliminate any instrumental modulation created by the rotating analyzer followed by a birefringent filter, which is very sensitive to polarization, the rotating quarter-wave plate would have to have the same precise specifications as the fixed quarter-wave plate. This modulation error produces a bias in the data which can be eliminated through a careful calibration.

The greatest difficulty in developing a vector magnetograph has been the accurate measurement of linear polarization to obtain the transverse component of the magnetic field. Instrumental polarization and circular cross talk have proven to be major problems in designing and building ground-based vector magnetographs. A vector magnetograph flown in space with a rotating analyzer as the polarimeter design will offer the solar community the opportunity to obtain very accurate polarization measurements, on the order of 1×10^{-4} .

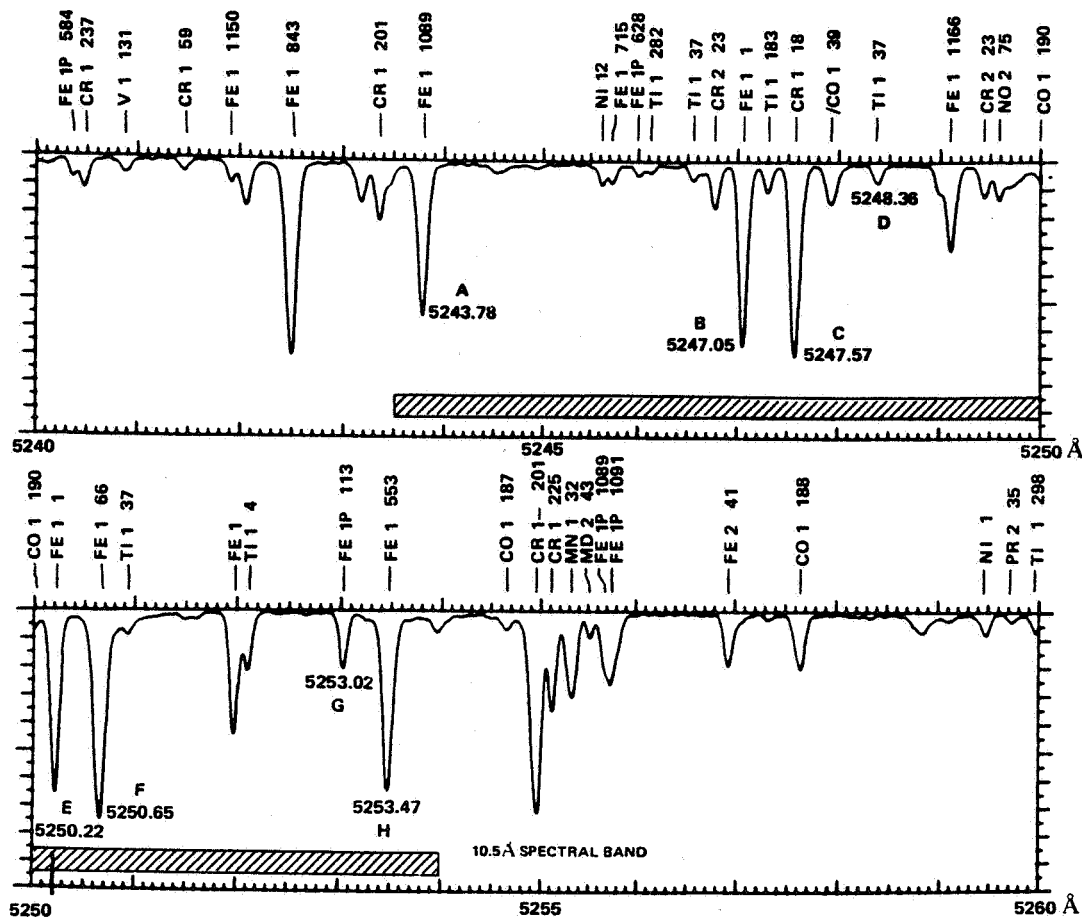
5. Tunable Spectral Filter

5.1 Introduction

A tunable spectral filter is required to isolate solar spectral lines that are magnetically sensitive (i.e., that exhibit the Zeeman effect), to extract a portion of a line profile for magnetic measurements, and to extract portions in the red and blue wings of a line profile for Doppler measurements. We have determined that an optimum filter for the SAMEX magnetograph should have a bandpass of 120 mÅ (full width at half maximum, FWHM), a tuning resolution of 1 mÅ, and a narrow, ~10 Å spectral range that covers spectral lines suitable for flux tube diagnostics (e.g., for deducing the pressure and temperature effects on the Zeeman lines).

The spectral range selected for the SAMEX magnetograph is from 5243.5 to 5254.0 Å, an interval covering 10.5 Å as shown in Figure 37. Although it covers only a limited portion of the visible spectrum, this wavelength range provides the spectral data needed to address all our scientific objectives through the use of analytical techniques for flux tube diagnostics. The major spectral lines in this range are Fe I 5247.05 (Landé g factor of 2.0), Fe I 5250.2 ($g = 3.0$), and Fe I 5250.6 ($g = 1.5$). These three lines have been used extensively in ground-based work, so the analysis techniques to obtain vector magnetograms of the photospheric fields, perform flux tube diagnostics, and derive the temperature dependency of the lines are well-developed (Stenflo, 1971, 1985). The 10.5 Å spectral range also allows measurement of the magnetically sensitive lines Fe I 5243.78 ($g = 1.5$), Cr I 5247.57 ($g = 2.5$), Fe I 5253.02 ($g = 1.0$), and Fe I 5253.47 ($g = 1.5$).

The limited spectral range of 10.5 Å is preferred over a range covering the full visible spectrum since it will minimize significantly the complexity of the magnetograph's design for the telescope, polarimeter, and filter. A critical component of the polarimetric measurement is the quarter-wave plate. By not having a broadband requirement placed on this achromatic wave plate, a very accurate quarter-wave plate can be fabricated for the 10.5 Å bandpass, and the accuracy of the polarimetric measurements will be correspondingly increased. A similar argument holds for the unique optical coatings used on the foreoptics to minimize instrumental polarization: the narrower spectral range reduces the design requirements for these coatings. The tunable spectral filter, if the birefringent type, is reduced in length compared with a filter that is tunable over the entire visible spectrum, and this shorter length increases the filter's étendue. Moreover, the filter's design is simplified since fewer and smaller optical components are needed. The cost, size, weight, and power requirements of the filter are thereby decreased and its expected lifetime on orbit should be increased. Finally, with the small bandpass that would be required for the prefilter, the solar intensity passed through



SELECTED SPECTRAL LINES:

<p>A: 5243.78 Fe I $5F_3 - 5F_4$ $G = 1.5$ $EP = 4.26$</p>	<p>C: 5247.57 Cr I $5p_1 - 5D_0$ $G = 2.5$ $EP = 0.96$</p>	<p>E: 5250.22 Fe I $5D_0 - 7D_1$ $G = 3.0$ $EP = 0.12$</p>	<p>G: 5253.02 Fe I $2p_2 - 5p_1$ $G = 1.0$ $EP = 2.28$</p>
<p>B: 5247.05 Fe I $5p_2 - 7D_3$ $G = 2.0$ $EP = 0.09$</p>	<p>D: 5248.36 Ti I $5F_1 - 5D_0$ $G = 0.0$ $EP = 0.81/1.88$</p>	<p>F: 5250.65 Fe I $5p_2 - 5p_3$ $G = 1.5$ $EP = 2.20$</p>	<p>H: 5253.47 Fe I $5D_1 - 5D_1$ $G = 1.5$ $EP = 3.28$</p>

Figure 37. The proposed spectral range for the SAMEX tunable filter. The eight major spectral lines in the 10.5 Å range from 5243.5 to 5254.0 Å are listed in the table below the spectra along with their atomic transitions, Landé g factors, and lower excitation potentials in eV. This spectral range covers several lines of Fe I with different excitation potentials and g factors. Hence we will be able to perform the flux tube diagnostics necessary to account for the effects of temperature, pressure, and filling factors.

the optical system to the filter is reduced, so that the overall internal thermal control requirements for the instrument are reduced.

To determine the best bandpass (FWHM) for a filter used in the SAMEX magnetograph, we have performed an analytical study of measured signal as a function of filter bandpass based on the characteristics of the Fe I 5250.22 line profile; this analysis is presented in section 5.4. Based on these results and other considerations, we believe that a filter bandpass of 120 mÅ represents a good trade-off between transmitted intensity and magnetic sensitivity.

To measure Doppler velocities, analyze line profiles (Stokesmeter mode), and correct for spacecraft heliocentric velocities, the filter must have a tuning resolution of approximately 1 mÅ. This specification is based primarily on the requirement for accurate calibrations of the Doppler measurements using the solar rotational velocity measured at the limb. The wavelength λ' of a spectral line emitted from a source moving with velocity v relative to an observer is shifted from the normal wavelength λ by an amount given by

$$\lambda' - \lambda = \Delta\lambda = \lambda \left[\frac{(1 - v/c)}{(1 - v^2/c^2)^{1/2}} - 1 \right].$$

Since $v/c \ll 1$, we can write $\Delta\lambda/\lambda \approx v/c$. The limb of the Sun has a radial velocity with respect to an observer on Earth of 2.1×10^3 m/sec. Thus, for this velocity, the relative wavelength shift $\Delta\lambda/\lambda$ is 7×10^{-6} ; this gives a wavelength shift of 37 mÅ at the wavelength 5250 Å. If the solar radial velocity is used to calibrate Doppler velocities measured with the SAMEX magnetograph, we must be able to measure the solar radial velocity to better than a few percent; this implies that the tuning resolution of the spectral filter should be on the order of 1 mÅ. For a tuning resolution of 1 mÅ, the resolution for Doppler velocities will be 60 m/sec.

A typical orbital velocity varies between ± 7 km/sec; this motion will produce periodic doppler shifts with peak amplitudes of 122 mÅ at 5250 Å. Thus the spectral filter may have to be shifted perhaps 16 times each 24-hour period to compensate for these Doppler shifts caused by spacecraft motion. The filter's tuning mechanism must be rugged enough to endure this constant, long-term frequency shifting so that there will be no degradation of filter performance in tuning.

Three different optical configurations for locating the tunable filter in the optical beam are in common usage: collimated, telecentric, and minimum volume. Our recommendation is to use the collimated configuration with the pupil located

near the center of the filter. This recommendation is based on the following analyses of each configuration and the effects of defects in a filter.

(1) Collimated configuration:

In a collimated configuration, the light from each object point is collimated through the filter so that the marginal rays are parallel to the optical axis. Light from different object points must pass through the filter at different angles. Since all the light from each object point is parallel, this configuration produces the minimum passband and highest resolution of all the configurations. However, the passband shifts with object coordinate since different light from different objects is collimated through the filter at different angles. Usually the collimated configuration is used with the pupil imaged into the center of the filter to minimize the size of the filter. In addition there will be a further wavelength shift due to solar rotation. Rust et al. (1986a) have shown how the wavelength shift with object coordinate in a Fabry-Perot filter can be used to partially counter the wavelength shift caused by solar rotation.

(2) Telecentric configuration:

In a telecentric configuration, the chief rays for all object points are made parallel to the optical axis by locating the pupil at infinity. Then the center wavelength in the passband is the same for all points in the field of view. The ray bundle from a given object point traverses the filter in a cone. The angular spread between the different rays from the same object point causes a loss of spectral resolution but eliminates the passband shift with object coordinate of the collimated configuration. Usually the telecentric configuration is used with the image located in the center of the filter to minimize the filter volume. To reduce possible image defects due to poor optical quality in the filter materials, it may be desirable to increase the size of the filter and move the image away from the filter.

(3) Minimum volume configuration:

The total volume of the filter can be minimized by locating equal sized images and pupils at opposite ends of the filter. This would minimize the amount of calcite needed in a birefringent filter. This configuration is "half way" between the telecentric and collimated configurations. There is a lesser wavelength shift with object coordinate than the collimated configuration and less passband broadening than the telecentric configuration.

Our choice of configuration is driven by the desire to reduce the impact of filter defects on the accuracy we wish to achieve in the polarimetric measurements (10^{-4}). Defects include

dirt, scratches, bubbles, crystal defects, and other scattering or absorbing objects in the beam. It is assumed that the tunable filter will contain more defects than the other elements that make up the optics of the SAMEX magnetograph. For example, birefringent filters are fabricated from calcite, a naturally occurring mineral. Large quantities of calcite are usually examined to find just a few pieces with minimum defects for constructing retarders, but there will always be some defects in the selected crystals. Moreover, birefringent filters used in both ground-based and spaceflight instruments have been known to develop bubbles. In Fabry-Perot filters, the light is reflected perhaps 10 times and therefore passes through the filter on the order of 20 times; consequently, the effects of any defects are magnified twenty-fold. These defects could include reflective coating defects, crystal defects, and microcracks which develop in the crystal due to the mechanical strain of modulation.

With these factors in mind, we must determine the best location of the tunable filter in the optical beam to minimize the impact of any defects on optical quality. Thus, the filter will be operated in a collimated configuration. This configuration provides the maximum immunity to optical defects located in the filter by distributing the effect of any one defect nearly equally over all image points. It also provides the highest resolution for a given filter, but at the expense of a field-dependent wavelength shift. However, we feel we can correct for this wavelength shift in the data reduction, whereas the degradation of polarimetric sensitivity by defects cannot be eliminated and would prevent our achieving the desired accuracy in measuring the vector magnetic field.

The tunable filter will be selected from one of two types available: birefringent and Fabry-Perot filters. Some basic characteristics of these two types of filters are listed in Table 43 (Rosenberg and Title, 1981; Rust et al., 1986a). In the next table, Table 44, we have listed certain specifications that are required for the spectral filter designed to be used in the SAMEX vector magnetograph.

In the following sections, we present overviews of both types of filter, discussing birefringent filters in section 5.2 and Fabry-Perot devices in section 5.3. In section 5.4, we compare their relative merits for application to high-resolution polarimetry. Finally, in section 5 we summarize the many considerations involved in selecting the optimum spectral filter for the SAMEX vector magnetograph.

Table 43. Basic characteristics of birefringent and Fabry-Perot filters.

Parameter	Birefringent Filters		Fabry-Perot Filters	
	Universal	Lyot	Solid etalon	Air-spaced
Type				
Source	Lockheed	Zeiss (MSFC)	Johns Hopkins University	Queensgate, Burleigh, and others
Aperture	35 mm	36 mm	62 mm	35 mm
Length	340 mm	256 mm	0.22 mm	30 mm
Optical	150	50	1	2 elements
Blocking filters	yes	yes	yes	yes
Control	computer, motors, thermal	motor, thermal	laser, voltage, thermal	pz device
Temperature sensitivity (A per °C)	0.5/quartz 0.3/calcite	0.5/quartz 0.3/calcite	0.04	
Wavefront	$\lambda/10$	$\lambda/10$	$\lambda/30$	$\lambda/100$
Spectral range	all visible	$\lambda \pm 8 \text{ \AA}$	$\lambda \pm 2.6 \text{ \AA}$	$\lambda \pm 1.8 \text{ \AA}$
FWHM	60 mÅ	120 mÅ	120 mÅ	100 mÅ
Transmission	20%	27%	53%	20%
Wavelength tuning accuracy	2 mÅ	10 mÅ	0.4 mÅ/volt 38 mÅ/°C	
Wavelength stability (mÅ per hr)	10		0.01	0.5
Finesse	37		22	15

Table 44. Optical specifications for the spectral filter for the SAMEX vector magnetograph.

Wavelength range	5243-5254 Å
Bandpass (FWHM)	120 mÅ
Tuning resolution	1 mÅ
Length of filter	<330 mm
Wavelength variation with angle of incidence	as small as possible

5.2 Birefringent Filters

Birefringent filters are interference devices that have been used in solar spectroscopy for many years. These filters are composed of a series of birefringent elements sandwiched between parallel polarizers; the thicknesses of successive birefringent elements are in a constant ratio of 2:1. The first polarizer of each stage converts the incident light into a single state of polarization. The birefringent element decomposes this light into two orthogonally polarized components and delays one component relative to the other. The exit polarizer forces the two components to recombine, and, in so doing, to interfere. The effect of each stage is to transmit some wavelengths and to reject others. By placing several stages in tandem, with the exit polarizer of one stage serving as the entrance polarizer for the next stage, the wavelength spacing between transmitted wavelengths - the free spectral range (FSR) - is increased since each smaller element causes the extinction of every other maximum of the previous stage. The thickest element determines the passband of the filter while the number of smaller stages determines the FSR.

5.2.1 A birefringent filter for the SAMEX vector magnetograph

In the following paragraphs, we will describe a particular birefringent filter designed by Lockheed Palo Alto Research Laboratory for this study. The description has been extracted from a report submitted to MSFC by Dr. W. J. Rosenberg of Lockheed; this report is reproduced in its entirety as Appendix A of this document. The Lockheed report is meant to be a discussion of the applicability of a calcite birefringent filter to a solar vector magnetograph. It is not a complete design nor a proposal to build such a filter.

5.2.2 Standard Lyot design

The standard birefringent filter is the Lyot design in which the filter consists of a set of crystal plates, each one-half as long as its predecessor (Lyot, 1944; Title and Rosenberg, 1981). Each plate is separated from its neighbors by efficient

polarizers having high transmission in the pass direction and high extinction (very low transmission) in the cross direction. Enough elements must be included in order to meet both the FWHM and FSR requirements. For the present application, a FWHM of 120 mÅ requires a longest element length of 60 mm. With seven elements, and consequently a shortest element length of 0.938 mm, the FSR is nearly 17 Å. Two plots of the resulting filter profile are shown in Figure 38. The first plot (a) shows a full free spectral range including two filter peaks. A detail of the peak showing the typical Lyot sidelobe structure is shown in the second plot (b).

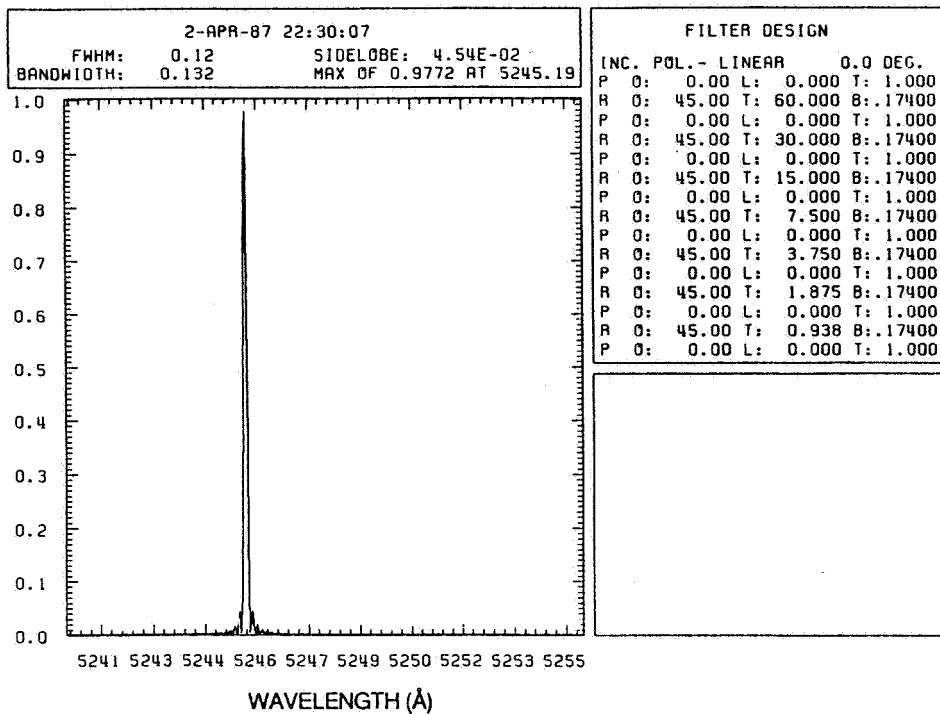
The specific plate thicknesses for this seven plate Lyot filter are:

Plate	Thickness (mm)
1	60.
2	30.
3	15.
4	7.5
5	3.75
6	1.875
7	0.938

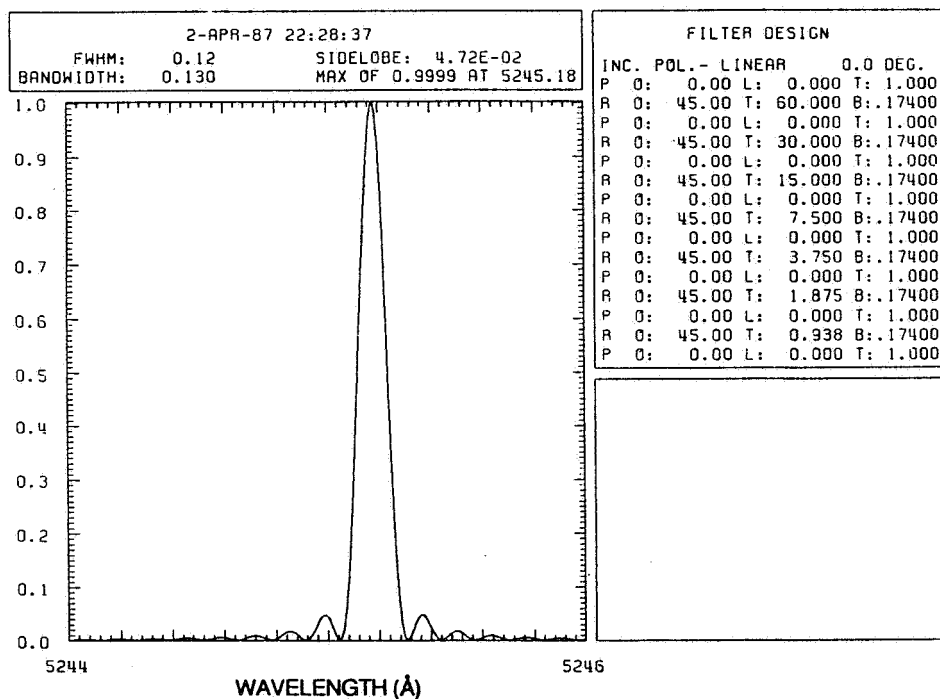
In practice, it is difficult to make a field widened calcite plate which is less than 1.0 mm thick so the seventh plate must be modified. The best approach is to use a three-halves plate instead of the one-half plate. In this example a three-halves plate is 2.813 mm thick. The resulting change to the transmission profile is shown in Figure 39. It is apparent that there is almost no change. The sidelobe structure is typical of a Lyot design with a maximum sidelobe level of about 5% of the mainlobe. For the design bandwidth (FWHM) of 120 mÅ the integrated noise bandwidth, or equivalent width, of the filter is 130 mÅ. Of that equivalent width, 95 mÅ, or 73%, falls between the half-maximum points. The ratio of the FWHM passband transmission to that of the wings is, therefore, 2.7.

5.2.3 Partial-polarizing configuration

In a manner mathematically equivalent to optical aperture apodization, the response profile of a birefringent filter can be modified to reduce the out-of-band sidelobes. While there are a number of techniques which effect this reduction, the most practical is the partial polarizing method (Title and Rosenberg, 1981; Rosenberg, 1986). The calcite elements are separated by leaky, partial polarizers rather than the high extinction polarizers of the Lyot design. This has two advantages: the sidelobes are reduced as desired, and the overall transmission of the filter increases. The transmission increase is due to the polarizers having better transmission in their pass direction as well as leaking light in their extinction direction. The tunable birefringent filter flown on Spacelab 2 (Title et al., 1986) is an example of a partial polarizing design.

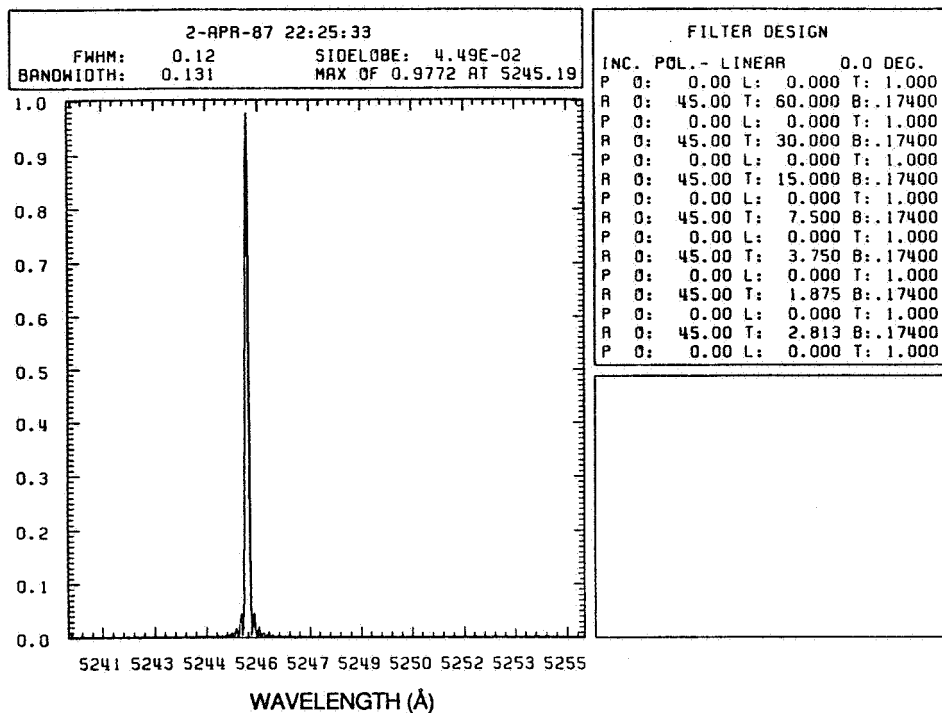


a

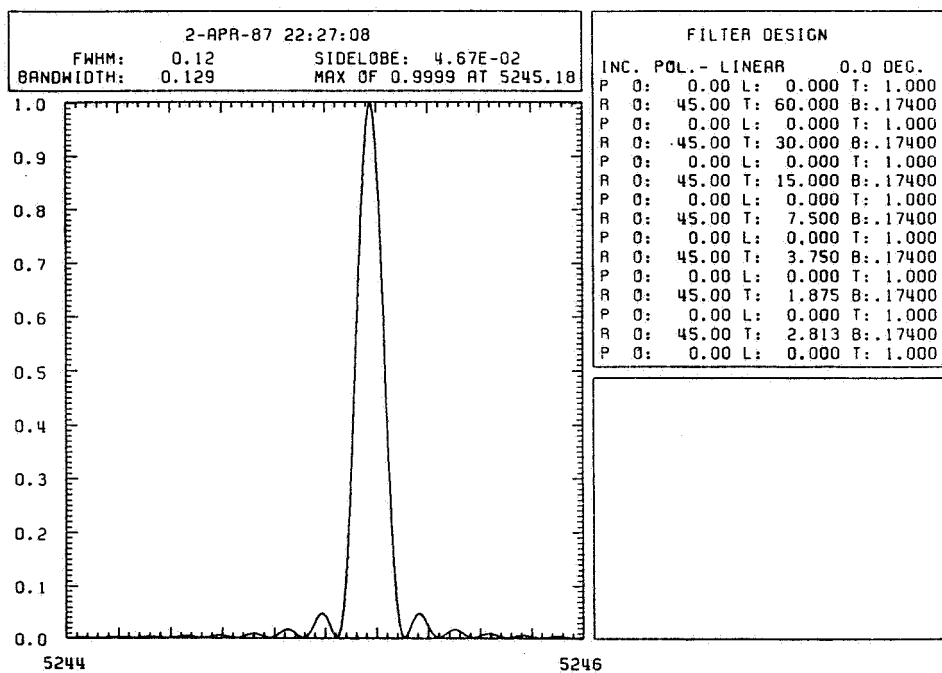


b

Figure 38. Transmission profiles of a standard, seven-element Lyot filter with a FWHM of 120 mÅ. (a) Transmission profile over the full free spectral range of the filter. The relative transmission of the filter is plotted as a function of wavelength in Å. (b) Transmission profile over 2 Å. The typical Lyot sidelobe structure is seen on this scale.



a



b

Figure 39. Transmission profiles of a modified, seven-element Lyot filter. The thinnest plate in the design of the filter of Figure 38 is difficult to fabricate (0.94 mm), so a 2.8 mm thick three-halves plate is used instead of the very thin one-half plate. This figure shows the transmission profile that results from this modification; it is apparent that there is almost no change. (a) Transmission profile over the full free spectral range. (b) Transmission profile over 2 Å.

The spectral profile of a candidate partial polarizing filter is shown in Figure 40. Again, the first graph (a) shows a full free spectral range of the spectrum, while the second (b) is a detail of the filter's passband. Notice that the out-of-band sidelobes have been reduced to less than 1% instead of the 5% Lyot sidelobe levels. For this partial polarizing design, the FWHM is 120 mÅ and the equivalent width is 126 mÅ. Of that equivalent width, 97 mÅ, or 77%, falls within the half-maximum points. The ratio of the FWHM passband transmission to that of the wings is 3.3.

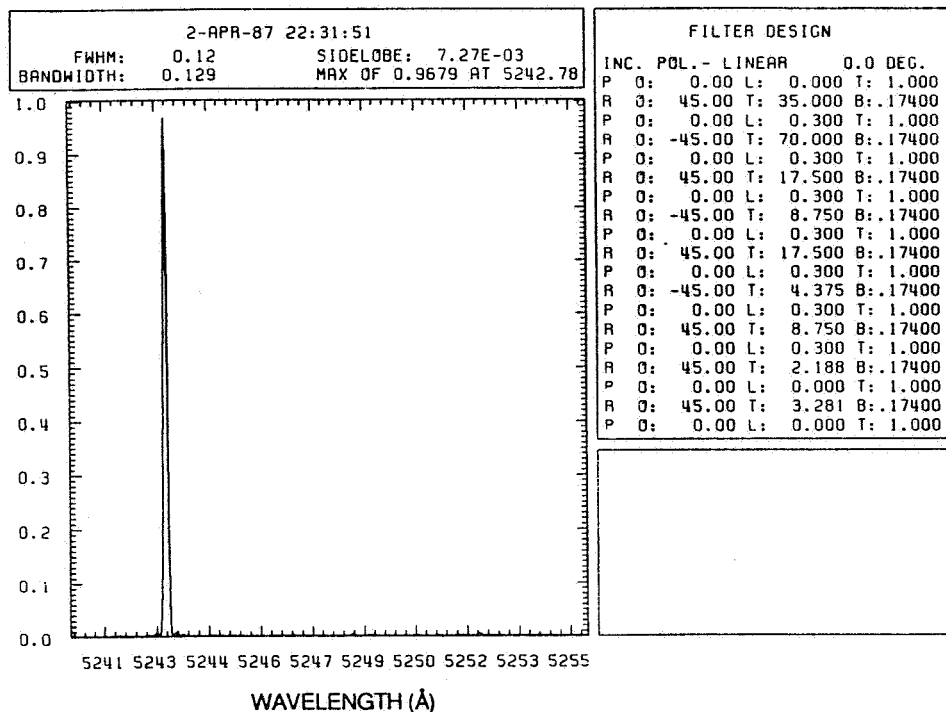
The specific plate thicknesses for this nine plate partial polarizing filter are:

Plate	Thickness (mm)
1	35.
2	70.
3	17.5
4	8.75
5	17.5
6	4.375
7	8.75
8	2.188
9	3.281

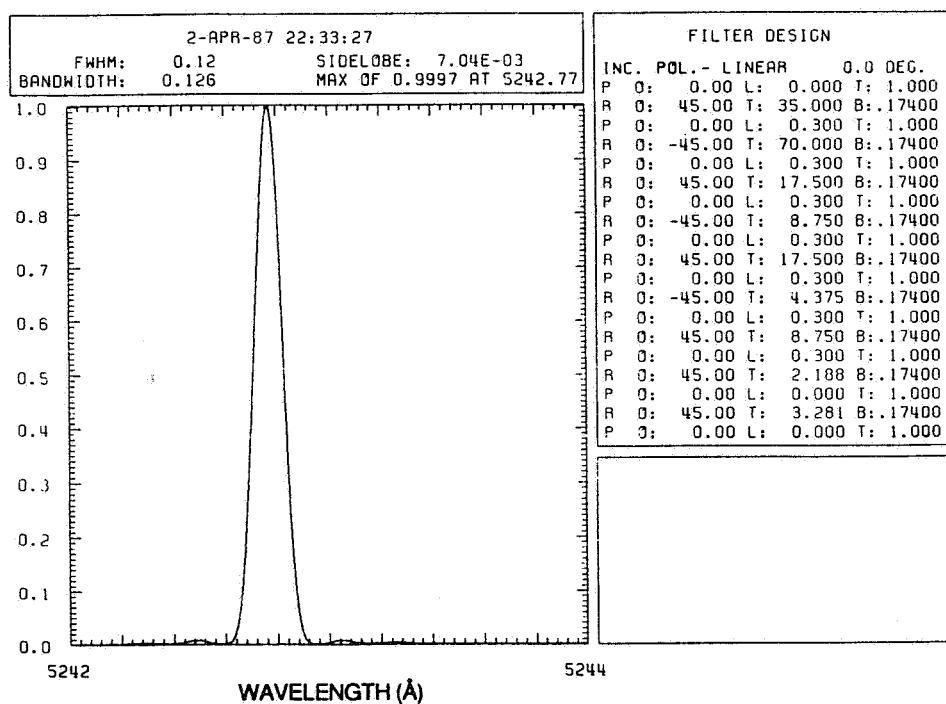
5.2.4 Tuning

One of the most remarkable properties of birefringent filters is the tuning mechanism (Title and Rosenberg, 1981). If a single element is considered, and the exit polarizer is replaced by a quarter-wave retardation plate, then the output of the element is linearly polarized light in which the orientation of the plane of polarization varies with color. A rotatable polarizer following the quarter-wave plate, therefore, becomes a tuner and selects the wavelength which corresponds to its orientation. Intuitively, it can be approached as follows. The output of a birefringent retardation plate whose input is linearly polarized light is elliptically polarized light. Fortunately, the eccentricity of the ellipse varies with wavelength while the orientation of the ellipse axes remains fixed, parallel to the input polarization. The subsequent quarter-wave plate converts elliptical polarization to linear polarization. The orientation of the polarization is dependent on the eccentricity of the ellipse, and therefore on the input wavelength. Tuning is accomplished by a mechanical rotation of a polarizer following the quarter-wave plate.

For Lyot filters with high extinction polarizers, the tuning technique works quite well and has been standard practice on birefringent filters for many years. Rotation of the exit polarizer by 180° tunes the element through a free spectral range of the respective element. For the longer, narrower bandwidth elements, the amount of rotation to effect a given wavelength shift becomes larger. The sensitivity to angular errors is



a



b

Figure 40. Transmission profiles of a partial-polarizing birefringent filter. In this design, the calcite elements are separated by leaky, partially polarizing polarizers instead of the high extinction polarizers used in the standard Lyot design. (a) Transmission profile over the full free spectral range. (b) Transmission profile over 2 Å. Here the sidelobes seen in the transmission profiles of the previous two designs are almost absent.

constant, independent of the spectral resolution of any given element. It is, therefore, no more difficult to tune a 0.1 Å bandwidth element as a 100.0 Å one.

Since the orthogonal polarization is absorbed by the rotating exit polarizer in a Lyot configuration, it has not been a concern in most filters. In a partial polarizing system, however, both polarizations are important and must be considered. As it turns out, there is one problem which must be corrected. The orthogonal polarization is one-quarter wave out of phase and will not interfere properly in the subsequent element. This condition is easily corrected by the addition of an achromatic quarter-wave retardation plate oriented parallel to either the transmission or extinction axes of the rotating polarizer. In previous partial polarizing filters using the alternating partial polarizing design, this was not necessary because each element has an efficient polarizer at one of its ends. For each element the efficient polarizer end was used for tuning. The all partial polarizing design of a quartz filter built at LMSC (Lockheed) in 1986 was the first filter which required the additional quarter-wave plate in its tuning section.

The tuning resolution depends on the angular resolution of the rotation mechanism. A rotation of the polarizer by 180° tunes an element by one free spectral range of that element. For the Lyot design listed above, the longest element is 60 mm and has a free spectral range of about 260 mÅ. In the partial polarizing design the longest element is 70 mm with a free spectral range of 225 mÅ. In previous filters Lockheed designers have used stepping motors to drive the tuning components. These have always had 48 steps per free spectral range which would correspond to a tuning resolution of 5.5 and 4.7 mÅ for the Lyot and partial polarizing designs, respectively.

There is no reason why greater angular resolution cannot be substituted for the 48 step per revolution motors. In plans for future filters, Lockheed intends to use brushless dc motors with absolute encoders for rotation control. With these the resolution is almost unlimited, being mainly a question of the precision of the encoder. Lockheed is specifying 4096 position encoders, not because that accuracy is required, but because those 12-bit encoders are convenient. With such a motor-encoder pair, the tuning resolution would be less than 0.1 mÅ. This would be misleading precision, however. With a temperature tolerance of roughly 300 mÅ/°C, a 1.0 mÅ tuning resolution would correspond to a temperature uniformity and stability of only a few milli-degrees. Similarly, the wavelength uniformity across the aperture is limited by the thickness uniformity of the calcite crystal element. The wavelength shift is approximately 7 mÅ/micron of thickness variation. With a bandwidth of about 120 mÅ, it is extremely difficult to measure thickness variations or wavelength shifts on the order of 1.0 mÅ. While the tuning mechanism could have the required precision, the tuning accuracy would in fact be limited by other factors.

5.2.5 Field of view

The feature that made birefringent filters useful to Lyot and subsequent solar physicists is the possibility of relatively large fields of view (Lyot, 1944; Title and Rosenberg, 1979). By constructing a wide field element, the spectral sensitivity to angle of incidence variations can be minimized. A wide field element is constructed by taking a simple crystal element and cutting it into two pieces, each half the length of the original. By placing these two identical plates at a relative orientation of 90° , the variation of total retardation with incident angle is minimized. A half-wave retardation plate (or a 90° rotator) separating the two plates is required to really make it work properly. The key to this construction is the azimuthal behavior of individual uniaxial crystal retardation plates. The retardation decreases with incident angle along the optic azimuth and increases on the orthogonal azimuth.

For field-widened calcite elements at an incident angle of ± 16.6 millirad, in air, there is a wavelength shift of $15 \text{ m}\text{\AA}$ relative to the on-axis spectral response. This shift is to the red and is nearly independent of azimuth. Since the intent is to place the filter in a collimated portion of the optical system, this will mean that there will be an apparent red shift of the filter toward the outside of the image plane. The outer pixels will see a $15 \text{ m}\text{\AA}$ redder portion of the spectrum than the on-axis pixels. If the filter were placed at an image plane, in a telecentric system, then there would be no spectral variation over the image. Instead, the apparent bandwidth of the filter would be broadened by the $15 \text{ m}\text{\AA}$, but would be constant over the image. These considerations are arguments for having a telecentric beam at the filter location in the SAMEX magnetograph's optical design.

5.2.6 Temperature effects

Calcite birefringent filters drift blue with temperature at a rate of roughly $300 \text{ m}\text{\AA}/^\circ\text{C}$. There are two problems which this causes. First is the change in wavelength of the filter as a whole as its temperature changes, and second is the non-uniformity of the spectral response caused by temperature variations across the aperture of the filter. In Lockheed calcite filters the attitude has been taken that since the filter is tunable, the temperature can be measured and the filter tuned to correct for any changes. With the filter in a sufficiently large aluminum case the thermal inertia makes this quite practical. The aluminum case also serves to minimize the presence of non-uniform temperature distributions at the periphery of the calcite elements. It does not, however, rule out the presence of radial temperature gradients in the crystals. Lockheed observations have shown that this is a surprisingly small effect. Measurements have been made of the spectral uniformity across filter apertures and no significant

temperature-induced non-uniformities have been found. For ambient temperature derivatives of several degrees C per hour, no spectral non-uniformities have been found at the 2 or 3 mÅ sensitivity of the measurements.

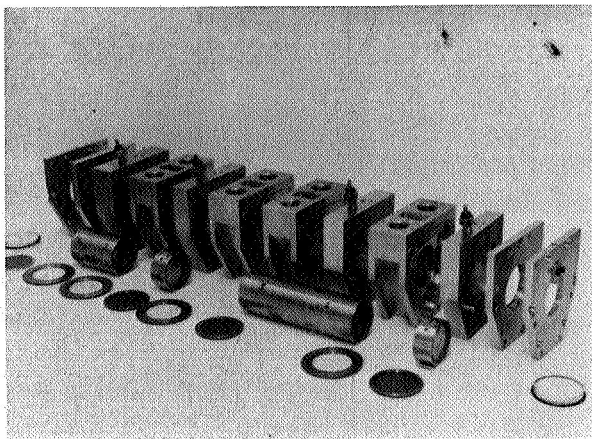
One technique which has been suggested for combatting the 300 mÅ per degree drift is to use a compensating crystal in conjunction with the calcite. A promising candidate would be ADP (ammonium dihydrogen phosphate) because of its relatively high retardation variation with temperature. At 5250 Å, ADP has a wavelength shift of almost 5 Å/°C. While this makes it difficult to use as a filter material by itself, it suggests using a relatively short piece to counteract the wavelength shift of calcite. This technique has not been tried by Lockheed. They have reservations about it because the sensitivity of the ADP might make localized temperature gradients more of a problem. In other words, by trying to fix an already manageable problem, the as yet non-existent problem might be promoted into real trouble.

5.2.7 Physical characteristics

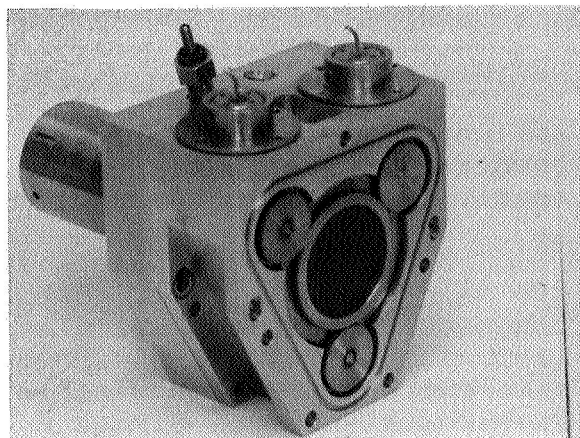
The birefringent filters built by Lockheed consist of the calcite elements contained in an aluminum package. The elements themselves sit in an oil-filled cavity which provides optical coupling to minimize reflection losses and ghosts, while allowing free rotation of the tuning components. The tuning motors couple to the rotating components through ferrofluidic seals to contain the oil. Previous filters have used 48 position stepper motors with home position sensors to maintain control of the tuning components' rotation. A brushless dc motor with high resolution absolute shaft angle encoders is now believed a superior choice. The Spacelab 2 flight filter package is typical of the physical configuration of a birefringent filter for this application. Its construction is illustrated in Figure 41.

One problem which has been present in the oil-filled tunable filters has been the appearance and subsequent continued presence of small bubbles in the optical path. It is believed that better filling techniques, currently being employed in the Spacelab 2 filter refurbishment, will mitigate against the formation of bubbles. A packaging redesign which minimizes small passages and closely spaced rotating components should facilitate the removal of any bubbles which do form.

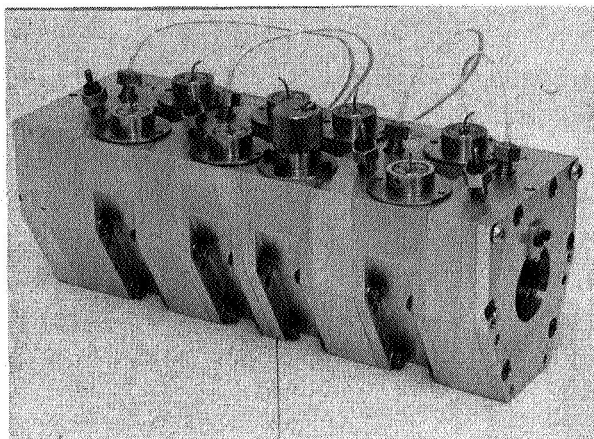
The question of clear aperture depends almost entirely on calcite availability. At the present time there appears to be a supply of calcite crystals consistent with a 25 mm clear aperture. There is the possibility, though not yet established, that larger aperture crystals might be available. Lockheed is thinking in terms of 50 mm to possibly 75 mm clear apertures. This possibility is being pursued independently of this application.



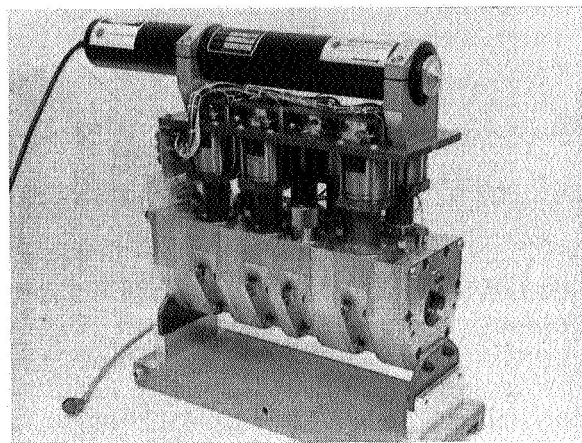
A



B



C



D

Figure 41. The Lockheed birefringent filter flown on Spacelab 2. This figure indicates the complex assembly of a flight qualified tunable filter. (a) Complete kit of optical and case components. (b) Assembly detail of single case section including calcite housing holding the two longest elements. (c) Completed assembly of optical and case components including rotary penetration joints for tuning motor coupling. (d) Complete filter including tuning motors and calibration laser.

ORIGINAL PAGE IS
OF POOR QUALITY

ORIGINAL PAGE IS
OF POOR QUALITY

The length of the filter depends on the choice of filter designs. For the Lyot design there is a total length of calcite of 121 mm. The seven elements require approximately 54 mm of waveplates, polarizers, and clearance. This implies a total optical length of about 175 mm, though it could be slightly more depending on the exact element packaging technique. The partial polarizing design uses about 170 mm of calcite, and requires about 70 mm of overhead for its nine elements. This implies a length estimate of 240 mm.

5.2.8 Transmission

The transmission of the Spacelab 2 flight filter at 5309 Å is 14% in polarized light. This is representative of what would be expected in this application. Since the filter need only tune over a limited range around 5250 Å, it should have somewhat better transmission. The polarizers can be optimized for the blue-green and the various wave plates can be simple monochromatic retarders instead of the achromatic laminations required for a universal filter. Since estimates of transmission often assume as much as 1% loss per layer of oriented plastic film, the use of single layer monochromatic retarders has a significant effect in a seven or nine element filter.

The birefringent filter would only be blocked over the 12 Å region around 5250 Å. In order to isolate that band a narrow passband, multilayer dielectric, interference filter will be required. The transmission spectrum of a typical filter is shown in Figure 42.

5.3 Fabry-Perot Filters

Fabry-Perot filters operate on the principle of an optically resonant cavity. The optical cavity is defined by two highly reflecting parallel surfaces separated by a distance d , while the cavity medium can be either air (air-spaced etalon) or a transparent material (solid etalon). Light that enters the cavity is partly transmitted and partly reflected. The reflected ray undergoes many reflections until its intensity is essentially extinguished, but it gives rise to a whole series of transmitted rays with path differences given by $2 n d \cos \theta$, where θ is the angle of incidence within the cavity medium of index of refraction n . The condition for a transmission maximum is

$$2 n d \cos \theta = m \lambda,$$

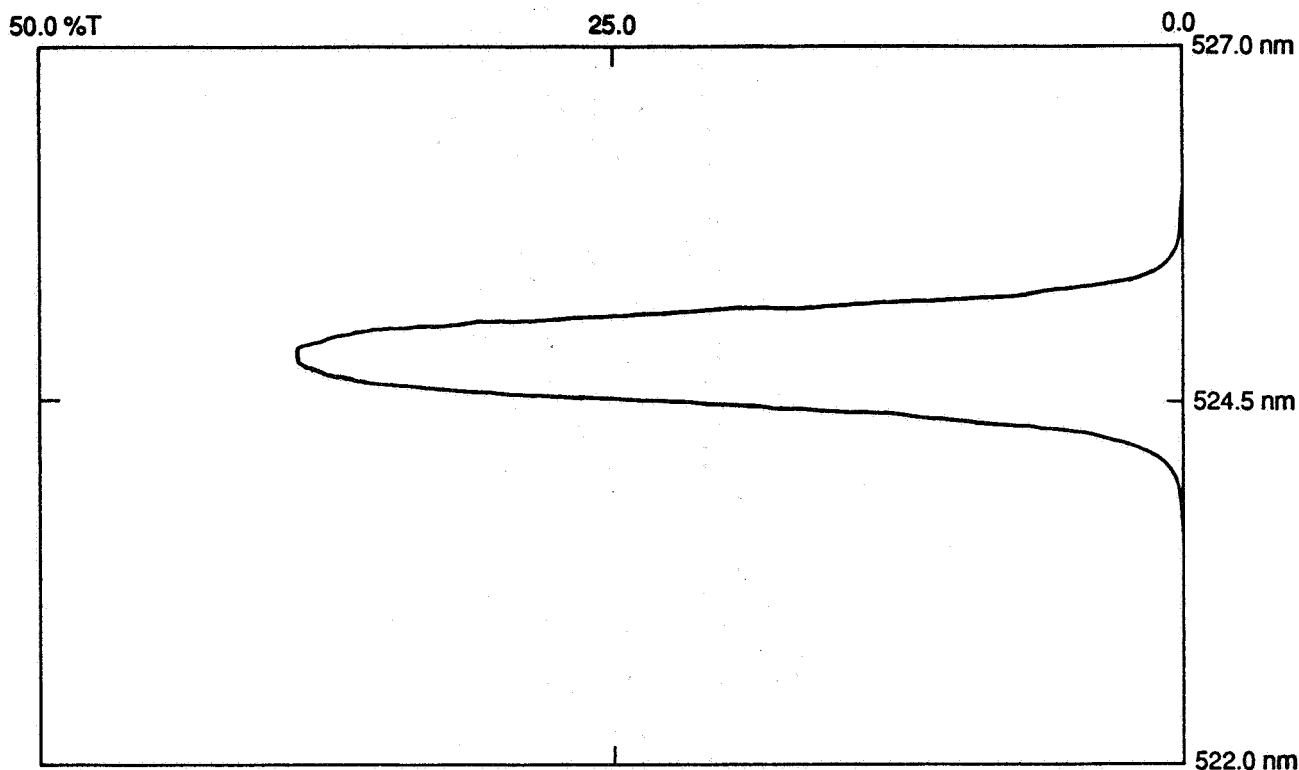
where m is an integer. This relation indicates that tuning, that is, changing λ , can be done by changing n , d , or θ . Thus, for air-spaced etalons, tuning can be achieved by varying the air pressure (changing n), tilting (changing θ), or varying the plate spacing d , usually with a piezoelectric transducer. In some solid etalons, n is changed by applying an electric field.

DATE: 03:26:87
CELL: _____

OPERATOR: _____
TEMPERATURE: 24° C
HUMIDITY: _____

SAMPLE: 5250 Å FLT
SOLVENT: BW 5A-7A
CONC: METHOD 55 USED

DATA MODE : (1) %T
BAND WIDTH : 0.10 nm
TIME CONST : 0.4 sec
λ SET : 527.0 ~ 522.0 nm
λ SCALE : 0.5 nm/cm
SCAN SPEED : 4 nm/min
%T ABS SCALE : 0.0 ~ 50.0 %T
CYCLE NO : 1



29.72 %T	525.0 nm
38.80 %T	524.8 nm
38.20 %T	524.9 nm

Figure 42. Transmission profile of a typical blocking filter. A birefringent filter transmits light at wavelength intervals equal to its free spectral range. To eliminate the transmission over all wavelengths except the desired one, a blocking filter must be used; the passband of the blocking filter should be centered at the wavelength of interest and have a passband slightly less than twice the free spectral range of the birefringent filter. This figure shows the transmission spectrum of a multilayer dielectric, blocking filter for the 5250 Å spectral range. This filter is a flight qualified, three period, interference filter that would be required as a blocker for the birefringent filters described in the text.

The filter transmits light of particular wavelengths, producing a comb-like spectral profile with peaks in transmission spaced in wavelength by the free spectral range (FSR) given by

$$\text{FSR} = \lambda^2 / (2 n d).$$

To make a good monochromatic filter, the FSR should be as large as possible to minimize the unwanted effects on the transmission profile of the blocking filters. These blocking filters are required to remove all but one of the peaks in the "comb." Since we have fixed the wavelength λ to be in the range 5243.5 to 5254.0 Å, only two parameters can be varied to maximize the FSR: the index n and the plate separation d . In solid etalons, the smallest attainable value of d depends upon the brittleness of the spacer material and the desired aperture A . Air-spaced etalons do not have this limitation, but for them, the index n is fixed (but small, i.e., equal to 1.00, compared to solid etalon indices, e.g., 2.3 for lithium niobate).

There is another factor in comparing the relative merits of air-spaced and solid etalons: the passband shift for off-axis rays, given by

$$\delta\lambda/\lambda = \phi^2 / (2 n^2),$$

where ϕ is the angle between the incoming ray and the normal to the filter. In this case, the larger index is definitely an advantage since it reduces the unwanted passband shift. The passband of a Fabry-Perot filter with a lithium niobate spacer will shift only about 11 mÅ at 5250 Å for rays at the solar limbs, whereas an air-spaced etalon will shift by about 57 mÅ. For this and other reasons, a solid etalon of this type is preferable over an air-spaced one for solar observations.

5.3.1 A Fabry-Perot filter for the SAMEX vector magnetograph

In the following paragraphs we will describe a particular Fabry-Perot filter - a high-throughput, tunable solid Fabry-Perot (FP) etalon - recently developed for the study of solar seismology by Rust et al. (1986a) at the Applied Physics Laboratory (APL) of the Johns Hopkins University. This description has been extracted from a report submitted to MSFC by Dr. D. M. Rust of APL; this report is reproduced in its entirety as Appendix B of this document.

5.3.2 Etalon design considerations

The performance of a solid FP etalon in an optical system depends on the index of refraction of the substrate, the free spectral range, the finesse, the diameter of the clear aperture, and placement relative to the pupil and image planes. The etalons under consideration here are constructed of lithium niobate, which is a highly transparent crystalline material whose index of refraction changes in proportion to the applied

voltage. The advantage of using lithium niobate crystal is that the FP passband can be tuned by the application of voltage to the crystal faces (Gunning, 1982). For lithium niobate, a voltage of ± 250 V will shift the etalon's passband by ± 0.1 Å.

In an FP, the higher the refractive index of the spacer, the less the passband shifts and broadens for off-axis rays since the wavelength shift is inversely proportional to the square of n , the index of refraction of the spacer. Thus, the acceptance angle of the lithium niobate etalon, for which $n = 2.3$, is 5.3 times that for an air-spaced etalon. Another way to express this advantage is to note that for the same spatial and spectral resolution, an air-spaced etalon must be 5 times the diameter of a lithium niobate etalon.

The FWHM of the passband is FSR/F , where F is the effective finesse, which depends upon the flatness, smoothness, and reflectivity of the coated surfaces. Figure 43 shows the transmission characteristics of a lithium niobate etalon made at the Commonwealth Scientific and Industrial Research Organization (Australia) Division of Applied Physics. The thickness of the etalon was 0.22 mm and the wavelength of the light was ~ 5700 Å.

Experience with a lithium niobate etalon (Rust, et al., 1986b) has established that a 50-mm etalon with $\text{FWHM} = 175$ mÅ is suitable for two-point observations of a Fraunhofer line profile in a low-resolution imaging system. In a study of possible etalons for the high-resolution SAMEX vector magnetograph, we have arrived at the following etalon specifications:

Filter passband (FWHM)	120 mÅ (at 5250 Å)
Finesse	22
FSR	2.6 Å (at 5250 Å)
Filter transmission	53%
Clear aperture	62 mm
Thickness	0.22 mm
Angle of incidence range in air	± 6.3 mrad

The operating wavelength and the FWHM were chosen on the basis of properties of the solar spectral lines and to maximize the signal-to-noise ratio. The FSR depends on the thickness of the etalon, as Figure 44 shows. The FSR should be as large as possible to ease the requirements on the auxiliary filters that will be needed to block all FP orders but one. However, it may be difficult to make a lithium niobate spacer much thinner than the 0.22 mm ones already in hand. Thus, we take $d = 0.22$ mm and $\text{FSR} = 2.6$ Å (ref. Fig. 2). Then, the required effective finesse (FSR/FWHM) is 22.

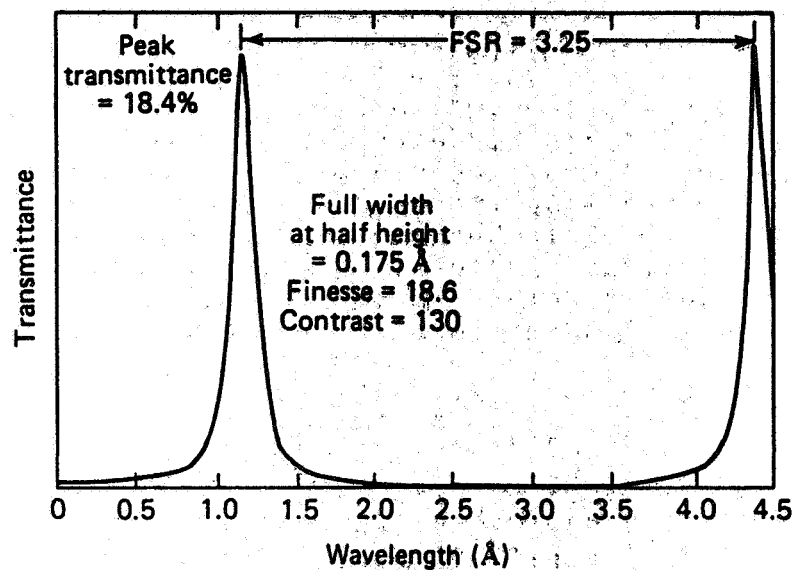


Figure 43. Transmission characteristics of a Fabry-Perot etalon. This figure shows the measured transmittance of a lithium niobate etalon fabricated at CSIRO in Australia.

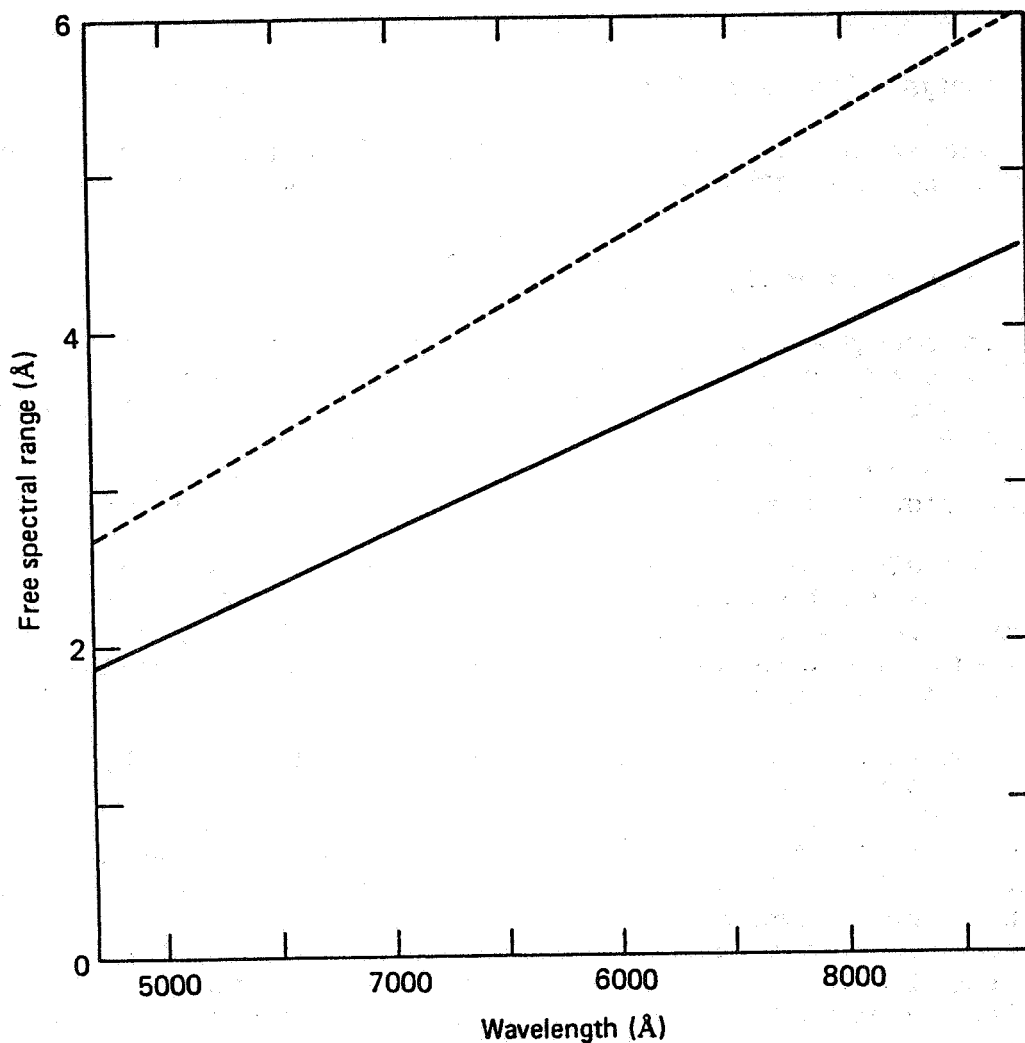


Figure 44. Variation of free spectral range with wavelength for a Fabry-Perot etalon. Shown is the free spectral range of a lithium niobate etalon as a function of wavelength for two thicknesses of the etalon: 250 μm (solid curve) and 175 μm (dashed curve).

Several factors contribute to the effective finesse, but for substrates polished to better than $\lambda/200$, the effective finesse is determined by the reflectivity R of the coatings:

$$F = \pi / \ln R ,$$

so, coatings with a reflectivity $>87\%$ are called for.

The maximum transmittance T_{\max} of an etalon is determined primarily by the reflectance R of the coating and its absorptance A :

$$T_{\max} = [1 - A/(1-R)]^2.$$

Silver is one practical choice for the coating material. For a 410 \AA layer of silver, $R = 0.87$ and $A = 0.035$ in the $5000\text{--}6000 \text{ \AA}$ range; hence T_{\max} is 53% . Multilayers of dielectric material (e.g., SiO_2 and TiO_2) are more durable than silver and one can achieve higher reflectance and thus greater finesse, although possibly at the expense of somewhat lower T_{\max} .

The clear aperture of 62 mm is near the largest practical size because lithium niobate is available in wafers only up to 75 mm diameter, and it is not always possible to achieve the degree of flatness and parallelism near the edges of the etalon that is required to give a finesse >20 there.

The range of the angle of incidence is found by dividing the system Lagrangian invariant of 0.195 mm radian by the semi-diameter of the etalon. The effect of a range of angles of incidence on the filtering properties depends on the optical system. In the telecentric system discussed below, the rays from a point in the image plane traverse the FP at the same angle, but this angle varies by ± 6.3 milliradians across the image. Therefore, for points at the extreme edges of the field of view, the filter passband will be offset by 20 m\AA from that seen by on-axis rays. There will be no additional broadening of the filter profile, however.

5.3.3 Tuning

The tuning requirements of a solar magnetograph are modest. Only Doppler shifts due to orbital motion of the spacecraft must be compensated for. These will be less than 0.2 \AA . To shift the passband from one wing of the spectral line to the other also requires $\sim 0.2 \text{ \AA}$ tunability. Thus, $\pm 1000 \text{ V}$ will provide adequate control over the passband. The tuning resolution or step size in wavelength can be of any magnitude from 0.4 \AA down to 4 \mu\AA , corresponding to voltage steps of 1000 V and 10 mV , respectively.

If the allowed voltage range is restricted to $\pm 1000 \text{ V}$, then only 30% ($0.8/2.6$) of the spectrum is accessible, and the lithium niobate FP cannot be considered a "universal filter."

However, the breakdown voltage of LiNbO_3 is much greater than 1000 V, probably over 4000 V (Gunning, 1982). After further testing, we expect that the operational voltages can be increased to 3000 V, at which point the filter will be able to tune to any wavelength, if provided with adequate blocking filters.

The FP may also be tuned by changing the temperature, if time is not critical. At 5250 Å, a temperature change of 20° K produces a wavelength shift of 0.76 Å. During solar observations, the etalon's temperature must be held constant to avoid distorting the magnetic field measurements. The temperature controller for an existing lithium niobate etalon holds the temperature constant to ± 0.01 °C. The induced uncertainty of 0.38 mÅ in the passband position is of no importance in solar magnetography. However, during an extended space mission, it is important to remeasure the central wavelength, voltage tuning parameter, and filter profile from time to time. For the lithium niobate FP, no special monitoring system is required for this because all critical filter parameters can be derived by tuning through a specified portion of the solar spectrum, which provides an adequate standard in itself.

The solar spectrum in one region of interest for magnetography is shown in Figure 37. Lines marked C and E are separated by 2.65 Å. With very skillful polishing techniques, an etalon may be worked to precisely that thickness which corresponds to a FSR = 2.65 Å. Thus, both lines may be sampled simultaneously by adjacent FP orders. This would allow the use of a 5 Å blocker to reject the unwanted orders rather than a 2.5 Å blocker, which would have a lower peak transmission. The simultaneous use of two lines would also improve the signal-to-noise ratio by a factor $\sqrt{2}$.

5.3.4 Line spread function

When used on solar spectral lines, the line spread function T of an FP is defined by the Airy function and takes the form

$$T(v_R) = \frac{1}{1 + \frac{4F^2}{\pi^2} \sin^2\left(\frac{\pi m v_R}{FR}\right)}.$$

In this expression, m is the order of the FP, v_R is the wavelength offset, expressed as velocity relative to the width Δv_0 of the spectral line, and FR is the "finesse" of the solar line, given by

$$FR = F m \Delta v_{FP} / \Delta v_0 = F m x,$$

where Δv_{FP} is the FWHM of the FP etalon. If

$$\pi v_R / FR \ll 1,$$

the function $T(v_R)$ can be approximated as

$$T(v_R) = \frac{1}{1 + 4(v_R/x)^2},$$

i.e., the profile is Lorentzian.

5.3.5 Blocking filters

To achieve its full promise, the FP filter should be provided with a tunable blocking filter, so that different orders can be selected quickly without having to rotate a filter wheel. One possibility for a tunable blocker is the acousto-optic filter (Harris and Wallace, 1969). Electronic tuning of the device is accomplished by tuning an acoustic wave generator.

Another possibility for a tunable blocker is the liquid-crystal filter described by Gunning et al. (1981). The properties of a prototype commercial liquid-crystal device (LCD) have already been measured (Rust, 1985). A modification of the traditional LCD design is needed to make a tunable blocker. A 25 μm layer of liquid crystal sandwiched between FP mirrors should produce a 1 \AA filter that can be tuned by application of a few volts to the mirror faces. The device is optically similar to the lithium niobate FP; its free spectral range is much higher, however, making it an efficient blocker for the lower FSR device.

5.3.6 Survival in space

An instrument in space is insulted by a variety of radiations and operational stresses that may limit its useful life. First is the vibration of launch, which conceivably could shatter the etalon, since the thickness of the lithium niobate substrate, which is more brittle than glass, is only 0.2 mm. APL intends to expose sample wafers of lithium niobate to simulated launch stresses. No difficulty is anticipated in space-qualifying an etalon, however, because standard techniques exist for protecting thin, brittle membranes during launch. The basic approach is to surround the membrane with acoustic-wave absorbing material.

The potential for radiation damage in lithium niobate depends on the orbit of the spacecraft. For equatorial orbits, one encounters energetic electrons, principally, and the etalon can be shielded from them. For a spacecraft outside the protecting envelope of the Earth's magnetic field, energetic protons could be a problem, since there is no practical shield against them. By disrupting the crystalline structure of the niobate, energetic protons could cause optical damage and lower

the transmission of the filter. However, experiments on lithium niobate samples (Allen, 1972) showed no change in transmission or index of refraction after exposure to 2×10^5 rads (AL) (1.8×10^{12} protons/cm² at $E = 85$ MeV, or 3.4×10^{12} protons/cm² at $E = 208$ MeV), which is an order of magnitude more radiation than expected on a 5-year mission at L1 with a minimum 1-mm-thick aluminum shield (JHU/APL, 1985). Lithium niobate was also exposed to 6.4×10^7 rads (AL) ($\sim 3 \times 10^{15}$ e⁻/cm²) in a beam of 2.2 MeV electrons with no change in transmission or refractive index. In these tests, the optical properties were measured at 5300 Å.

Damage to lithium niobate caused by optical beams has been studied extensively (Glass et al., 1972) and could be a problem in an etalon exposed to direct sunlight for an extended period. In a spaceborne magnetograph, prefilters will remove the damaging UV and most of the unwanted light. Solar optical radiation at ~ 5000 Å amounts to 0.02 mW/cm² Å, so with a telescope collecting area of 640 cm², a passband of 100 Å, and transmission efficiency of $\sim 25\%$ up to the FP (whose area is 30 cm²), the optical flux on the FP will be 10 mW/cm². Since the first mirror surface will reflect 90% of this, the flux inside the etalon would be only 1 mW/cm² were it not for the fact that the internal reflections amplify the beam there by a factor of F . With the etalon described above, $F = 25$, so the internal optical flux is 25 mW/cm². This is still much lower than the thresholds (300 W/cm²) at which optical damage has been observed with laser beam illumination.

In order to avoid distortion of the optics and to subject the etalon to a minimum thermal load, infrared radiation will be rejected by a prefilter over the face of the telescope. Most of the 30 mW of solar light passing through the first mirror surface is either transmitted by the etalon or reflected. Very little energy is absorbed, even by the silver coatings. Absorption can be further minimized by using thin film dielectrics for the reflective coatings. These films absorb less energy than metallic films and their long-term durability has been established (Title et al., 1974).

5.4 Comparison of Birefringent and Fabry-Perot Filters

There does not seem to be a clear-cut choice between the two types of filters we have considered - each has its own particular advantages and disadvantages. To provide an overview of these, we have compared the two filters with respect to some of the parameters important to the successful operation of the SAMEX magnetograph; these comparisons are summarized in Table 45. Parameters marked with an asterisk are those listed in Table 44, the primary specifications for the SAMEX filter.

Table 45. Comparison of birefringent and Fabry-Perot filters.

Parameter	Requirement	Birefringent	Fabry-Perot
Type		Lyot	Solid etalon
Source		Lockheed, Zeiss, etc.	APL
Aperture		25 mm	62 mm
Length*	<330 mm	240 mm	0.22 mm
Optical elements		~50	1
Blocking filters		yes	yes
Control		motors, thermal	laser, voltage, thermal
Temperature sensitivity (Å per °C)		0.3/calcite	0.04
Wavefront		$\lambda/10$	$\lambda/30$
Spectral* range (FSR)	11 Å	17 Å	2.6 Å
FWHM*	120 mÅ	120 mÅ	120 mÅ
Transmission		14%	53%
Wavelength* tuning accuracy (a)	1 mÅ	3 mÅ / 0.01 °C	0.4 mÅ / volt 0.4 mÅ / 0.01 °C
Wavelength* variation over FOV	minimal	15.4 mÅ	22.4 mÅ
Wavelength stability (mÅ per hr)			0.01
Finesse			22
Polarization sensitivity	minimal	yes	yes

(a) Assume temperature control to ± 0.01 °C

An important consideration in selecting the type of filter to be used in the SAMEX magnetograph is the magnitude of the linearly and circularly polarized intensities that are transmitted by the filter to the faceplate of the detector; several characteristics of the filter's profile influence this a great deal. One important characteristic in this respect is the filter's FWHM. If the FWHM is very broad, the filter integrates over oppositely polarized intensities of the different Zeeman components with the result that the measured intensity of a Stokes parameter is reduced significantly. For example, in the measurement of circular polarization, the V and I Stokes components are integrated over the transmission profile of the filter. Hence, the measured fractional circular polarization, P_V , will be given by

$$P_V = \bar{V} / \bar{I},$$

where the symbol $\bar{}$ indicates the integrated component. Since the left and right circularly polarized sigma components are displaced equally on either side of the central wavelength, the filter's transmission peak must be tuned to either the red or blue wing to maximize the measured signal P_V . With an increasing halfwidth this signal becomes more and more diluted as the transmission profile picks up more and more of the opposite circular polarization, decreasing \bar{V} , the net V signal. The same situation occurs in the linear measurement P_L , where

$$P_L = (\bar{Q}^2 + \bar{U}^2)^{1/2} / \bar{I}.$$

In this case the transmission profile integrates the two identically polarized sigma components on either side of line center with the oppositely polarized component at line center. And, again, a broader filter bandwidth reduces the contrast by reducing the net linear signal.

This contrast is also affected by the shape of the profile. If the profile has broad wings - that is, if the transmission $T(\lambda)$ doesn't fall off very rapidly with wavelength λ , or if it has strong secondary peaks at wavelengths away from the central transmission peak, then a significant contribution of unpolarized intensity from the nearby continuum may be integrated into the signal. Or the broad wings or secondary peaks may transmit parasitic polarization from neighboring spectral lines that are also Zeeman sensitive. In both cases, the contrast is reduced.

To examine the effects of these two filter characteristics - bandwidth and profile shape - on the signals P_V and P_L transmitted by a filter, we performed an analytical study, calculating P_V and P_L for three transmission profiles: a 1/8 Å bandwidth Fabry-Perot filter (1/8 FP), a 1/8 Å birefringent filter (1/8 BR), and a 1/16 Å birefringent filter (1/16 BR). We used line profiles generated by the Unno analytic equations (see Appendix D) for the Fe I 5250.22 spectral line. We studied two

different aspects of the problem: (1) the responses P_V and P_L as a filter is tuned through the line profile for a given field strength B and inclination (to the line-of-sight) ψ , and (2) the responses P_V and P_L versus field strength B for a given inclination ψ and wavelength.

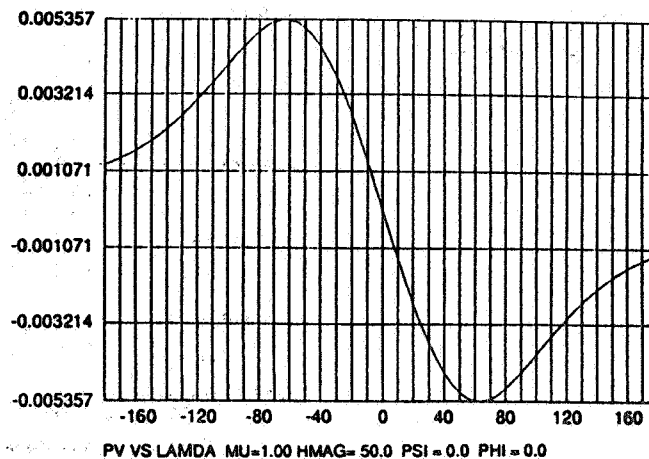
The responses P_V were plotted versus the wavelength of the peak transmission of the filter profile as the filter was tuned through the line profile for three values of field strength and an inclination of 0° ; these results are shown in Figures 45-47 for the three types of filters studied. Similar plots were generated for the response P_L for a completely transverse magnetic field ($\psi = 90^\circ$); these are shown in Figures 48-50. In Table 46 the maximum values of P_V and P_L are given for $\psi = 0^\circ$ and 90° for the three types of filters at different field strengths B .

Table 46. Maximum responses P_V and P_L for a $1/8 \text{ \AA}$ Fabry-Perot filter ($1/8 \text{ FP}$), a $1/8 \text{ \AA}$ birefringent filter ($1/8 \text{ BR}$), and a $1/16 \text{ \AA}$ birefringent filter ($1/16 \text{ BR}$).

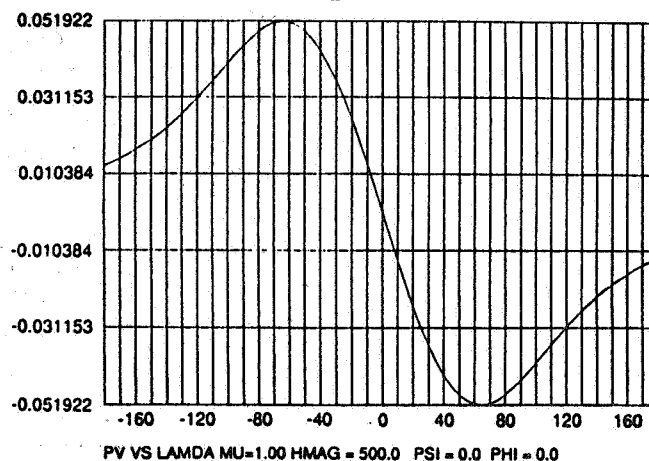
B (G)	$\psi = 0^\circ$			$\psi = 90^\circ$		
	1/8 FP	1/8 BR	1/16 BR	1/8 FP	1/8 BR	1/16 BR
50	.00536	.01021	.01827	.00022	.00030	.00051
500	.05192	.09859	.17056	.01804	.02351	.04206
1000	.09483	.17765	.28270	.04787	.06315	.11662
1500	.12505	.22790	.32823	.06976	.10074	.18236
2000	.14417	.25098	.34158	.08296	.13429	.22825
2500	.15588	.25825	.34730	.09327	.16097	.29307
3000	.16322	.26187	.35045	.10219	.20136	.32044

These data show that the $1/16 \text{ \AA}$ bandwidth filter gives the best response for both the linear and circular measurements. However, the curves in Figures 48-50 reveal a very interesting characteristic of these narrow bandwidth filters, one that has operational implications. For small field strengths, $<500 \text{ G}$, the linear signal (the one that gives the transverse field) for the $1/16 \text{ \AA}$ filter varies rapidly with wavelength over a very narrow wavelength range. This feature is in contrast to the broad wavelength range over which the $1/8 \text{ \AA}$ birefringent filter maintains a maximum linear signal. We have referred to this characteristic previously and indicated that it places much more stringent requirements on tuning the filter to the wavelength of

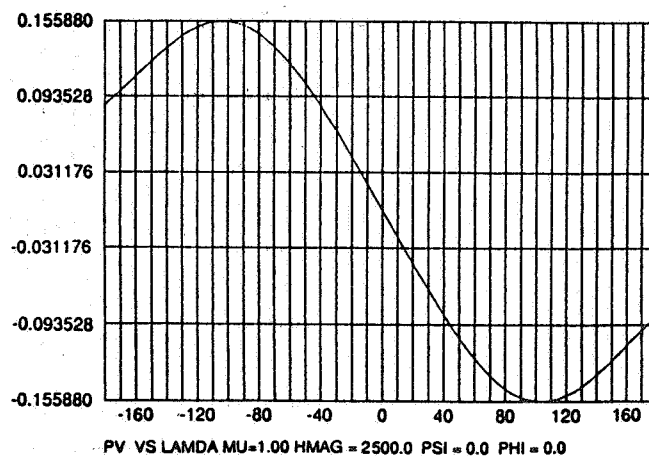
CIRCULAR POLARIZATION
UNNO THEORY WITH LINEAR SOURCE FUNCTION
BETA0 = 2.00 ETA0 = 10.00 DOPWD = 35.00
DAMPING CONST. = 0.10
FABRY-PEROT FILTER
1/8 A BANDPASS



a



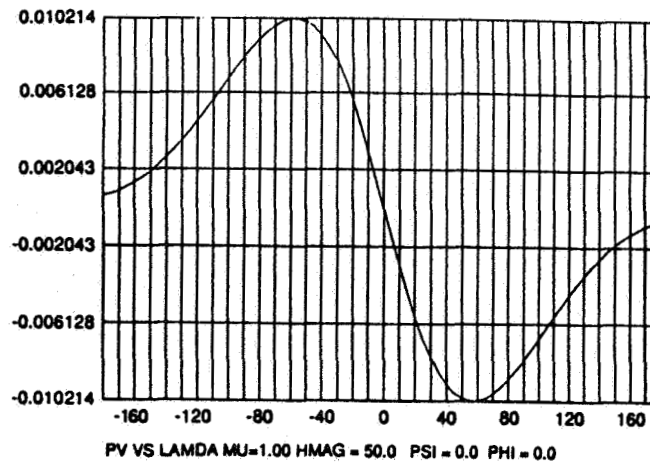
b



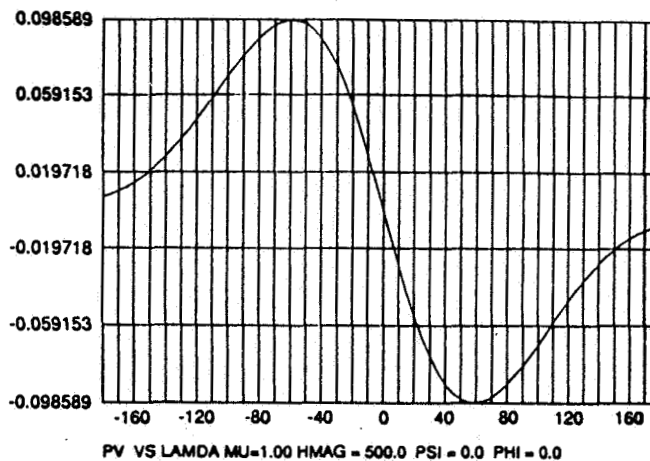
c

Figure 45. Theoretical variation of circular polarization transmitted by a 1/8 Å bandpass Fabry-Perot filter as the filter is tuned through a Zeeman-sensitive absorption line. The fractional circularly polarized intensity is plotted as a function of the wavelength offset in mÅ from line center (0 mÅ) for the case of a magnetic field along the line of sight ($\psi = 0^\circ$). The line formation parameters used in the analytical formulation of Unno (1956) are given at the top of the figure. (a) Magnetic field strength = 50 G. (b) Magnetic field strength = 500 G. (c) Magnetic field strength = 2500 G. Note that the peak signal scales almost linearly with field strength for the 50 and 500 G fields, but not for the strong field case.

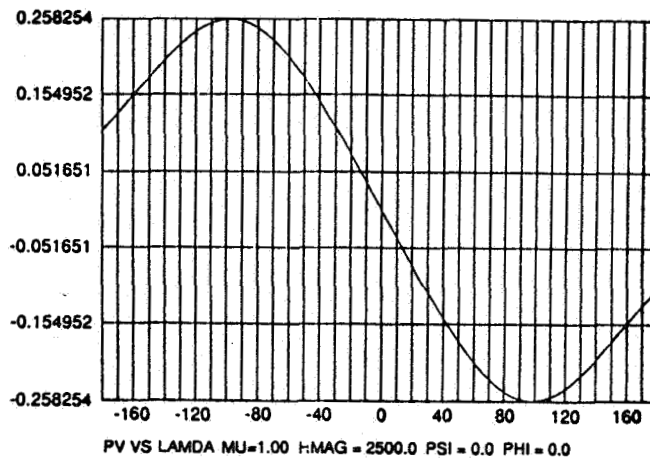
CIRCULAR POLARIZATION
 UNNO THEORY WITH LINEAR SOURCE FUNCTION
 BETA0 = 2.00 ETA0 = 10.00 DOPWD = 35.00
 DAMPING CONST. = 0.10
 BIREFRINGENT FILTER
 1/8 A BANDPASS



a



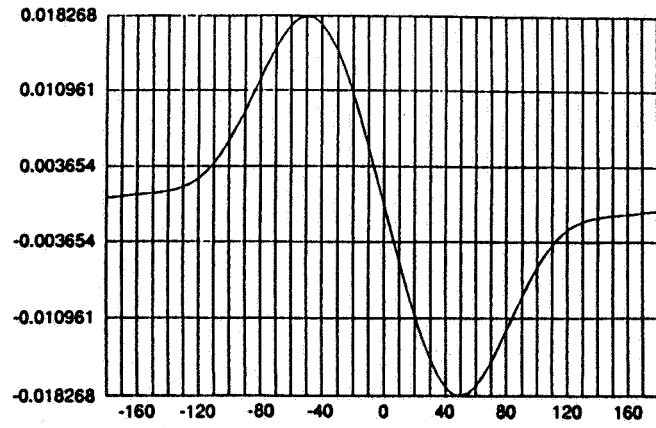
b



c

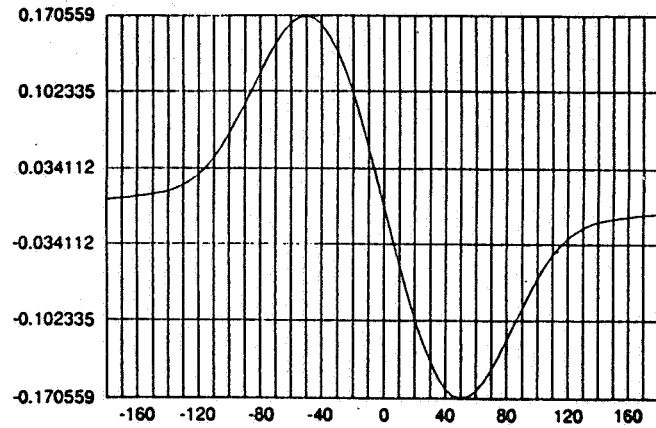
Figure 46. Theoretical variation of circular polarization transmitted by a 1/8 A bandpass birefringent filter as the filter is tuned through the line profile of Figure 45. All three panels represent the same cases shown in the previous figure.

CIRCULAR POLARIZATION
 UNNO THEORY WITH LINEAR SOURCE FUNCTION
 BETAO = 2.00 ETAO = 10.00 DOPWD = 35.00
 DAMPING CONST. = 0.10
 BIREFRINGENT FILTER
 1/16 A BANDPASS



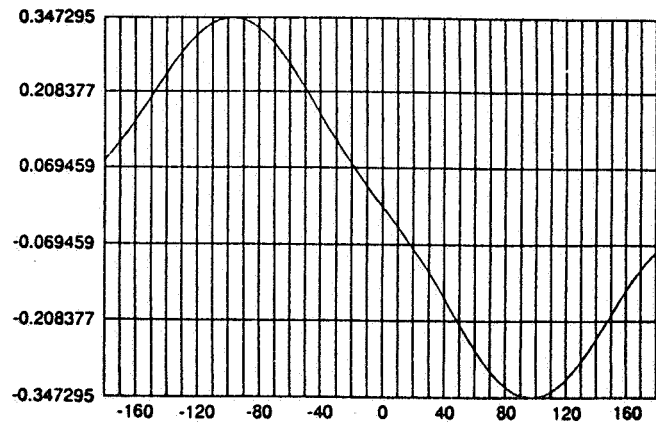
PV VS LAMDA MU=1.00 HMAG = 50.0 PSI = 0.0 PHI = 0.0

a



PV VS LAMDA MU=1.00 HMAG = 500.0 PSI = 0.0 PHI = 0.0

b



PV VS LAMDA MU=1.00 HMAG = 2500.0 PSI = 0.0 PHI = 0.0

c

Figure 47. Theoretical variation of circular polarization transmitted by a 1/16 A bandpass birefringent filter as the filter is tuned through the line profile of Figures 45 and 46. All three panels represent the same profile as shown in the previous two figures.

LINEAR POLARIZATION
 UNNO THEORY WITH LINEAR SOURCE FUNCTION
 BETAO = 2.00 ETAO = 10.00 DOPWD = 35.00
 DAMPING CONST. = 0.10
 FABRY-PEROT FILTER
 1/8 Å BANDPASS

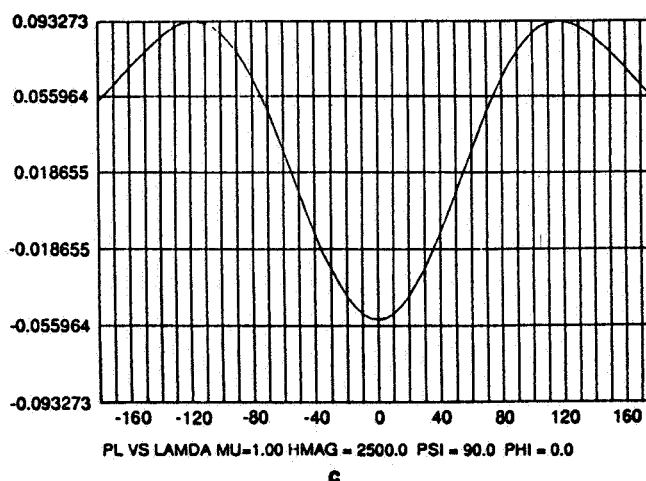
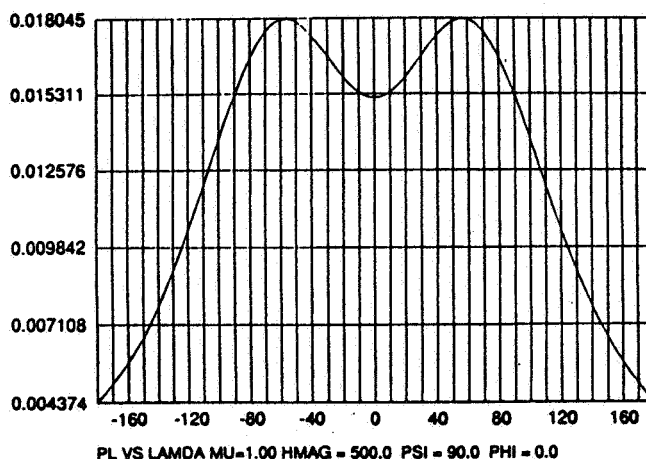
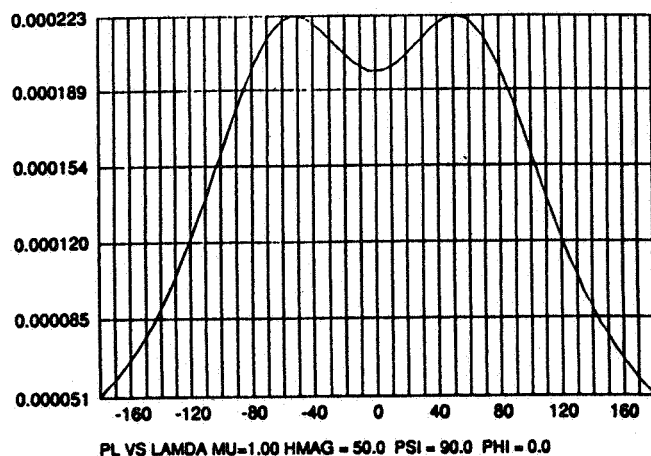


Figure 48. Theoretical variation of linear polarization transmitted by a 1/8 Å bandpass Fabry-Perot filter as the filter is tuned through a Zeeman-sensitive absorption line. The fractional linearly polarized intensity is plotted as a function of the wavelength offset in mÅ from line center (0 mÅ) for the case of a magnetic field transverse to the line of sight ($\psi = 90^\circ$). The line formation parameters used in the analytical formulation of Unno (1956) are given at the top of the figure. (a) Magnetic field strength = 50 G. (b) Magnetic field strength = 500 G. (c) Magnetic field strength = 2500 G. In the case of linear polarization, the peak signal scales quadratically with field strength for the 50 and 500 G fields.

LINEAR POLARIZATION
 UNNO THEORY WITH LINEAR SOURCE FUNCTION
 BETA0 = 2.00 ETA0 = 10.00 DOPWD = 35.00
 DAMPING CONST. = 0.10
 BIREFRINGENT FILTER
 1/8 Å BANDPASS

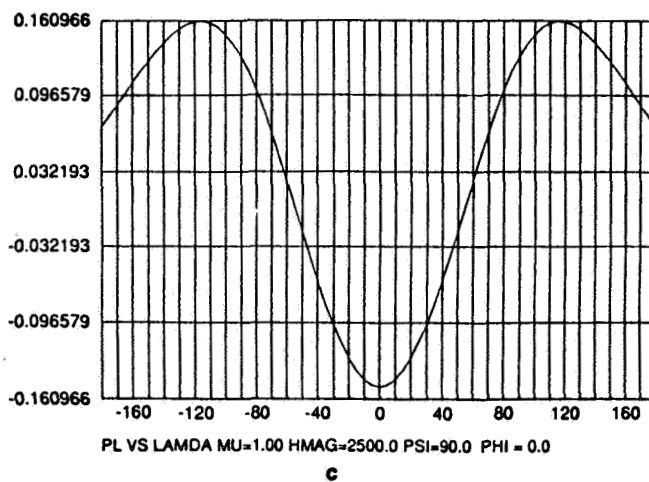
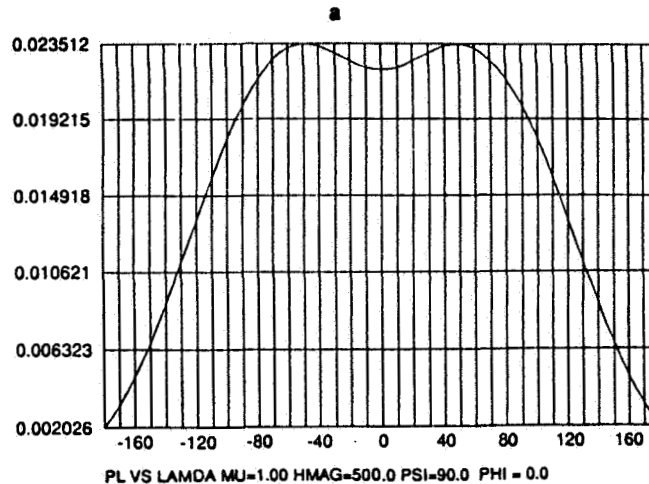
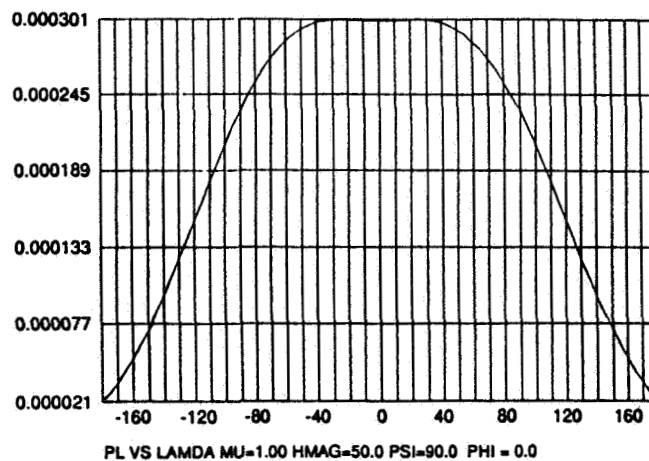


Figure 49. Theoretical variation of linear polarization transmitted by a 1/8 Å bandpass birefringent filter as the filter is tuned through the line profile of Figure 48. All three panels represent the same cases shown in the previous figure.

LINEAR POLARIZATION
 UNNO THEORY WITH LINEAR SOURCE FUNCTION
 BETA0 = 2.00 ETA0 = 10.00 DOPWD = 35.00
 DAMPING CONST. = 0.10
 BIREFRINGENT FILTER
 1/16 Å BANDPASS

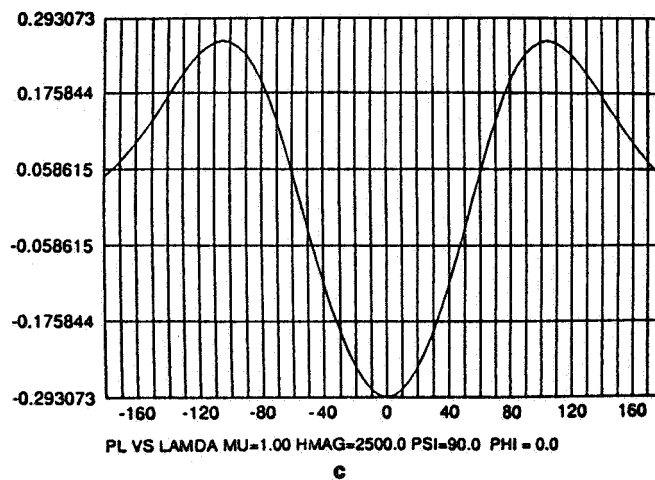
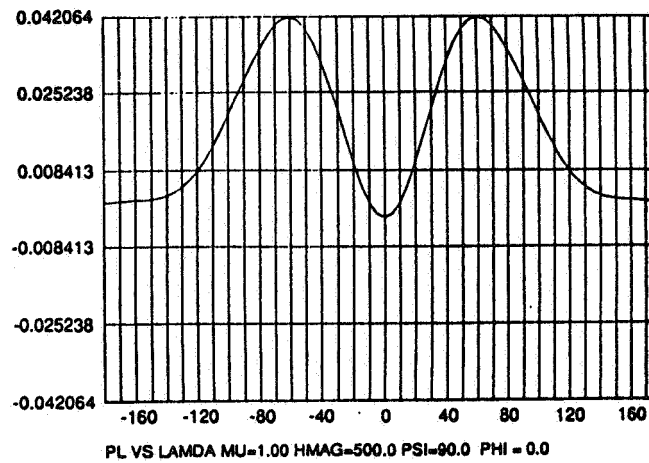
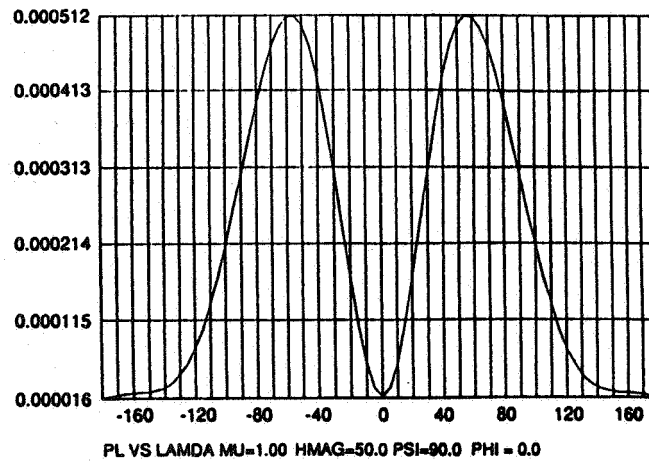


Figure 50. Theoretical variation of linear polarization transmitted by a 1/16 Å bandpass birefringent filter as the filter is tuned through the line profile of Figures 48 and 49. All three panels represent the same cases shown in the previous two figures.

maximum response and keeping the filter positioned at that wavelength. There is more room for error in this respect with the 1/8 Å filters. This feature offsets the advantage of the greater signal response of the narrower filter.

To get a sense of the linearity of the measured signals P_V and P_L as functions of field strength and to estimate the magnetic sensitivities for the three types of filters, we have plotted P_V and P_L versus field strength at the two inclinations $\psi = 0^\circ$ and 90° for the three filters. These plots are shown in Figures 51-53.

With regard to the range of linearity for P_V and P_L , the curves for the three filters are quite similar with deviations from linear relations occurring at just about the same field strengths for all three filters, namely at ~1000 G for P_V and ~500 G for P_L . Consequently, as far as achieving linear relations over a broader range of field strengths, there is no advantage in selecting one type of filter over the others.

These data afford us an estimation of the magnetic sensitivity that we might expect from measurements using these three types of filters, based on an expected polarimetric sensitivity of 10^{-4} and a signal-to-noise of 10^4 . These estimations are listed in Table 47.

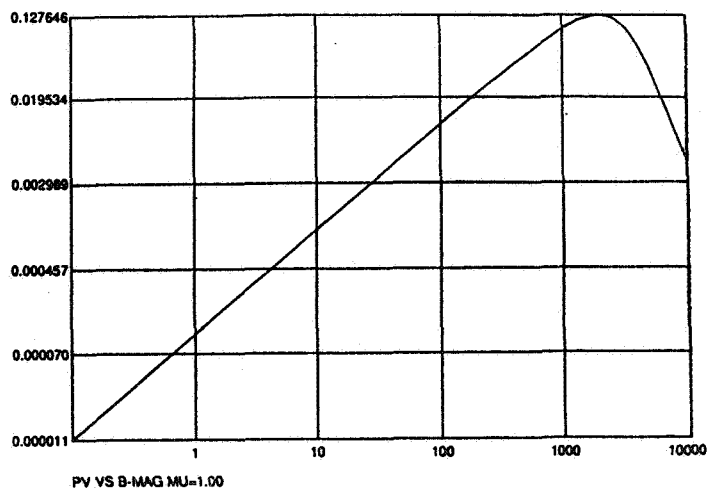
Table 47. Expected magnetic sensitivities for a 1/8 Å Fabry-Perot filter (1/8 FP), a 1/8 Å birefringent filter (1/8 BR), and a 1/16 Å birefringent filter (1/16 BR).

	<u>1/8 FP</u>	<u>1/8 BR</u>	<u>1/16 BR</u>
Minimum B_L	0.9 G	0.5 G	0.3 G
Minimum B_T	33.6 G	28.8 G	22.1 G

These estimations seem to indicate that all three filters would yield about the same level of magnetic sensitivity, so that the greater signal response of the narrower, 1/16 Å bandpass filter does not seem to be a significant factor in the decision concerning the type filter we would select for the SAMEX vector magnetograph.

Upon considering the results of this analytical study of signal response for different types of filters, we have concluded that the only significant differentiating factor is the broad wavelength range of maximum linear signal for the 1/8 Å

CIRCULAR POLARIZATION
UNNO THEORY WITH LINEAR SOURCE FUNCTION
BETAO = 2.00 ETAO = 10.00 DOPWD = 35.00
DAMPING CONST. = 0.10
FABRY-PEROT FILTER
1/8 Å BANDPASS SET AT -60.0mÅ
FIELD INCLINATION = 0.0



LINEAR POLARIZATION
UNNO THEORY WITH LINEAR SOURCE FUNCTION
BETAO = 2.00 ETAO = 10.00 DOPWD = 35.00
DAMPING CONST. = 0.10
FABRY-PEROT FILTER
1/8 Å BANDPASS SET AT 0.0mÅ
FIELD INCLINATION = 90.0

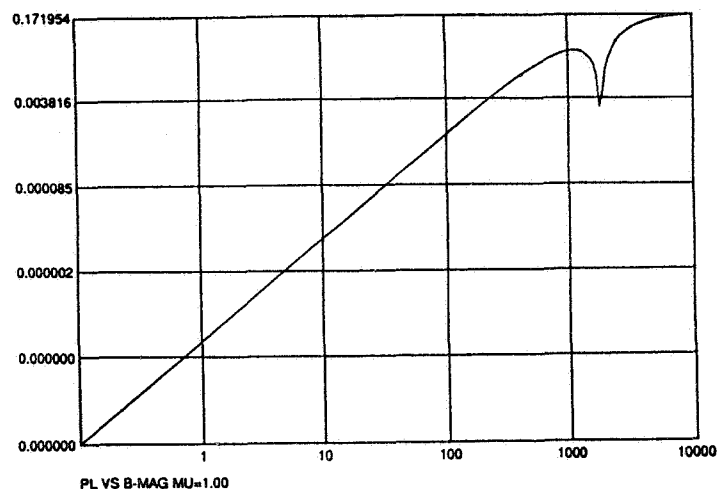
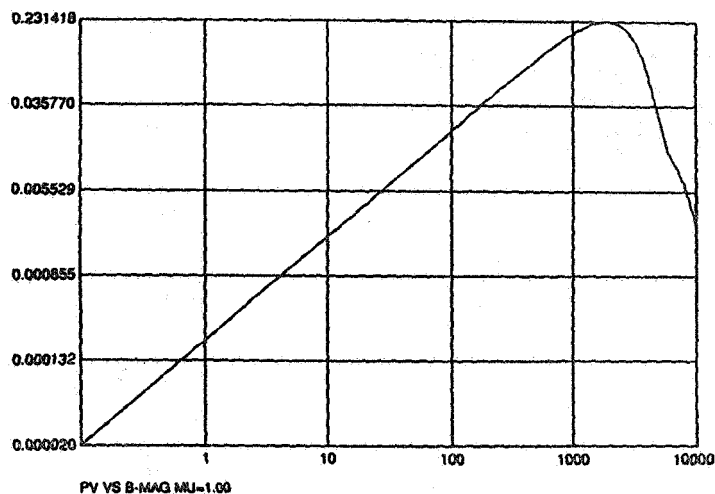


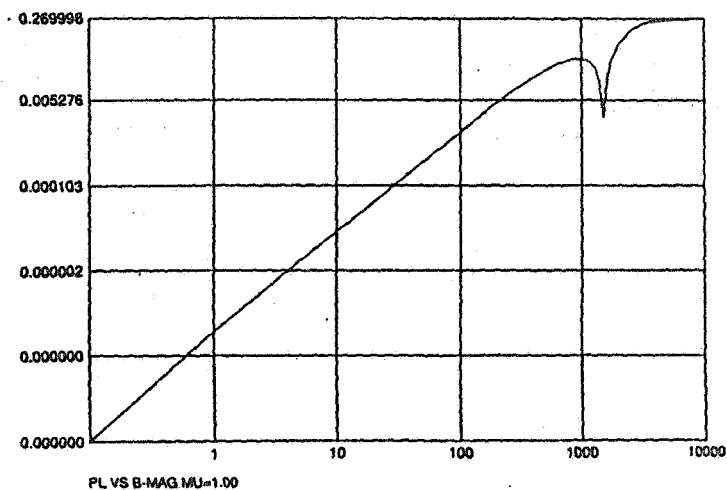
Figure 51. Theoretical variations of circular and linear polarizations transmitted by a 1/8 Å bandpass Fabry-Perot filter as a function of field strength. The line formation parameters used in the analytical formulation of Unno (1956) are given at the top of the figure. (a) Transmitted circular polarization for a magnetic field oriented along the line of sight ($\psi = 0^\circ$). Fractional circularly polarized intensity is plotted versus the strength of the magnetic field. The peak transmission of the filter is centered at 60 mÅ in the blue wing of the Zeeman-sensitive spectral line where the polarization signal is a maximum for weak fields (see Figure 45). (b) Transmitted linear polarization for a transverse magnetic field ($\psi = 90^\circ$). Fractional linearly polarized intensity is plotted as a function of the strength of the magnetic field. The filter is now centered at line center where the linear response is slightly less than optimum (see Figure 48). Note that the regime of linear response extends to about 1000 G for circular polarization, but, for linearly polarized light, the linear relation (on the log-log plot) fails well below 1000 G.

CIRCULAR POLARIZATION
 UNNO THEORY WITH LINEAR SOURCE FUNCTION
 BETA0 = 2.00 ETA0 = 10.00 DOPWD = 35.00
 DAMPING CONST. = 0.10
 BIREFRINGENT FILTER
 SYMMETRIC FILTER
 1/8 A BANDPASS SET AT -60.0mA
 FIELD INCLINATION = 0.0



a

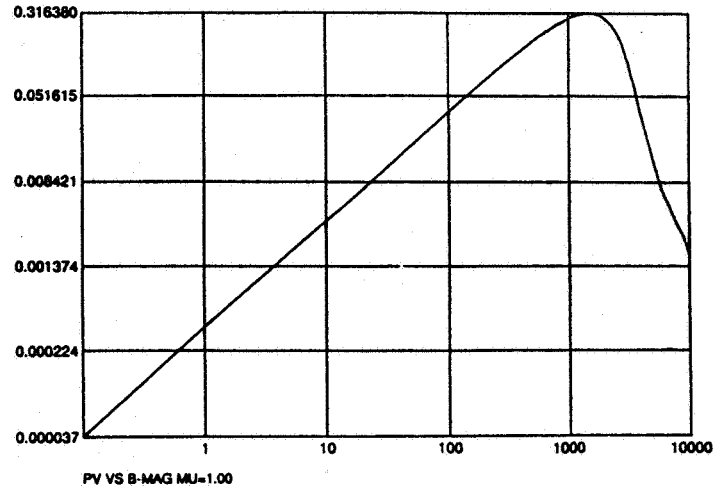
LINEAR POLARIZATION
 UNNO THEORY WITH LINEAR SOURCE FUNCTION
 BETA0 = 2.00 ETA0 = 10.00 DOPWD = 35.00
 DAMPING CONST. = 0.10
 BIREFRINGENT FILTER
 SYMMETRIC FILTER
 1/8 A BANDPASS SET AT 0.0mA
 FIELD INCLINATION = 90.0



b

Figure 52. Theoretical variations of circular and linear polarizations transmitted by a 1/8 A bandpass birefringent filter as a function of field strength. The two cases shown correspond to those in Figure 51. For this filter, the maximum response in the circular polarization is greater than in the case of the Fabry-Perot filter. The data given in Table 46 indicate the maximum responses in the linear polarization for these two filters is about the same for weak fields.

CIRCULAR POLARIZATION
 UNNO THEORY WITH LINEAR SOURCE FUNCTION
 BETAO = 2.00 ETA0 = 10.00 DOPWD = 35.00
 DAMPING CONST. = 0.10
 BIREFRINGENT FILTER
 SYMMETRIC FILTER
 1/16 A BANDPASS SET AT -50.0mA
 FIELD INCLINATION = 0.0



LINEAR POLARIZATION
 UNNO THEORY WITH LINEAR SOURCE FUNCTION
 BETAO = 2.00 ETA0 = 10.00 DOPWD = 35.00
 DAMPING CONST. = 0.10
 BIREFRINGENT FILTER
 SYMMETRIC FILTER
 1/16 A BANDPASS SET AT -60.0mA
 FIELD INCLINATION = 90.0

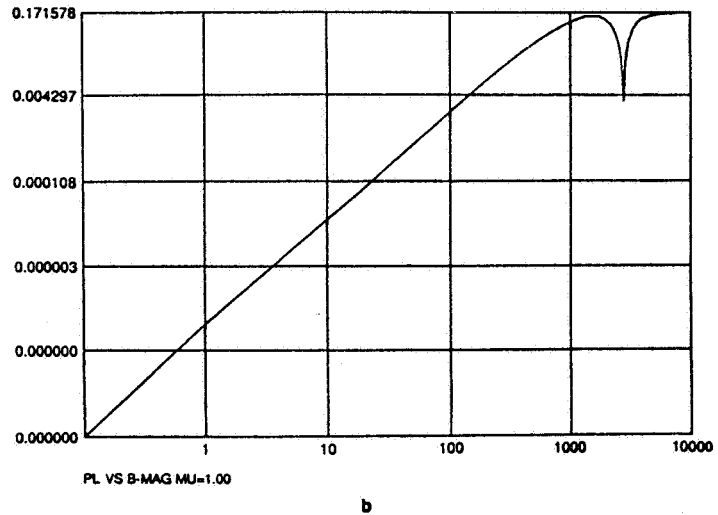


Figure 53. Theoretical variations of circular and linear polarizations transmitted by a 1/16 Å bandpass birefringent filter as a function of field strength. The two cases shown are similar to those in Figures 51 and 52, but the filter is centered at 50 mÅ in the blue wing for case (a) and at 60 mÅ for case (b) (see Figures 47 and 50). The maximum circular response with this filter is somewhat greater than for the 1/8 Å birefringent filter. The linear response for this filter is greater than for the two 1/8 Å bandpass filters (see Table 46).

C-3

birefringent filter for small field intensities (≤ 500 G). This factor is important because it places less stringent requirements on positioning the filter's peak transmission within this wavelength range and keeping it there.

5.5 Summary

It would seem that either the Lockheed or APL filter could be used in the SAMEX vector magnetograph. Considering the primary requirements listed in Table 44, both filters meet specifications on passband and length. The APL filter is superior in tuning resolution, but the Lockheed filter has the necessary FSR using just one blocking filter. Both suffer from variations in the wavelength of the passband over the field of view - a serious drawback. However, as indicated in the Lockheed report, this problem can be eliminated in a telecentric optical configuration. Thus, further consideration must be given to the relative advantages of the telecentric system to eliminate wavelength variations and the collimated system to minimize effects of optical defects in the filter. The bottom line in this choice is which system will give the greatest polarimetric accuracy.

In the optical design, we used optical parameters of aperture and length that are relevant to a birefringent-type filter. This choice was made only to provide the data needed to carry out a detailed optical design. It would be very easy to tailor the present design to accommodate a solid-etalon Fabry-Perot filter.

6. Detector System

In section 6 we review the characteristics of solid-state detectors. Those familiar with this material can skip to section 6.6.4 which describes the system proposed for SAMEX.

6.1 Introduction

The detector system must have the desired field of view, spatial and temporal resolutions, spectral range, and magnetic sensitivity to satisfy the scientific objectives. To cover the relatively large field of view of 4.3×8.5 arc min with a spatial resolution of 0.5 arc sec, allowing 2 detector pixels to resolve a 0.5" element, requires a detector with approximately 1000×2000 discrete imaging pixels. The requirement on magnetic sensitivity means that a signal-to-noise ratio (S/N) of 10^4 must be realized. The best S/N ratio for existing large-area detector systems has been only about 10^3 , so a process of enhancing the image by adding successive polarization images will have to be used (this process is used successfully by the MSFC vector magnetograph). Our experience has shown that as many as 225 arrays of 3 linear polarization images and 50 arrays of 2 circular images have to be accumulated to produce one vector magnetogram with a S/N of 10^4 . To meet the desired temporal resolution of ~ 5 minutes per vector magnetogram means that the total time to expose and read out each polarization image must be only a fraction of a second.

There are several types of detectors which could satisfy the requirements of the SAMEX vector magnetograph. However, our discussion will be limited to solid-state sensors since they are widely used in astronomy, are suitable for space since they are compact, low power devices, and their technologies have been extensively developed. There are two types of solid-state detectors that we have considered for the SAMEX vector magnetograph: charge coupled devices (CCDs) and charge injection devices (CIDs). The discussion on the detector system is limited to what we believe will be available in the next few years based on existing CID/CCD technology.

In the following sections, we will discuss and compare the operational characteristics of CCDs and CIDs and present our reasons for selecting a CCD camera system. We conclude with a description of the specific characteristics of a CCD camera system that will satisfy the scientific objectives of the mission while not incurring extensive customized development which would adversely impact the cost of SAMEX.

6.2 Design Specifications

The main scientific goals of the flight magnetograph are to obtain magnetograms with high spatial resolution (0.5 arc sec) and good time resolution (5 minutes) with a polarization accuracy of 10^{-4} . The specification for spatial resolution which

corresponds to 0.25 arc sec pixels has to be considered together with the requirements of the field of view. A necessary boundary condition on the observations is that the magnetic field of an observed active region be contained within the boundaries of the field of view. Since 90% of all active regions are contained within an area of 4×8 arc min², this establishes the dimensions of the focal plane array at approximately 960×1920 pixels. The optimum design would use a single detector although a focal plane array made up of multiple detectors is possible. The second major requirement that must be satisfied is the temporal resolution. To study the evolution of the magnetic field we have established a baseline of a 5-minute cadence between full vector magnetograms. Each vector magnetogram will contain five separate measurements of the Stokes vector. Ideally only four measurements are necessary to determine the Stokes vector. Our experience with the MSFC vector magnetograph has shown that changes between measurements require some redundancy. For example, one observing program to decrease the data acquisition time while obtaining equivalent magnetic field strengths between the longitudinal and transverse components would be to vary the enhancements between the Stokes polarization measurements: 50 enhancements for S_{v1} and S_{v2} , 225 enhancements for S_{q1} , S_{q2} , and S_{u1} (see section 2.4). Therefore each measurement must be completed in 60 sec and have adequate signal to noise to meet the requirement on polarization resolution (1 part in 10^4). The temporal resolution in turn sets limits on the exposure and the readout time of the camera system while the polarization resolution affects the signal-to-noise ratio (S/N).

The scientific requirements that affect the detector design are summarized in Table 48.

Table 48. Summary of scientific requirements on camera system.

Scientific Requirement	Detector Parameter
Spatial resolution: 0.5 arc sec	Array size & exposure time
Field of view: 4×8 arc min	Array size
Magnetic resolution	
Polarization accuracy: 10^{-4}	Signal-to-noise
Filter bandpass: 120 mÅ	Exposure time
Time resolution: ≤ 5 minutes (for a vector magnetogram)	Exposure & readout times

In order to compare the temporal resolution that can be achieved by different solid-state sensor designs, we must derive how many photons are available on a 0.25 arc sec pixel (the image size to be adjusted by the relay optics following the polarimeter). This is determined by the light collecting area of the telescope (which also affects the spatial resolution, see section II.1.3), the transmission characteristics of the optics (see section II.1.2), and the solar irradiance of the spectral line. Let $N(\lambda)$ be the number of photons per second incident on each 0.25" pixel with an average energy $E = hc/\lambda$ within an effective bandpass of $d\lambda$. Then

$$N(\lambda) = F A a d\lambda \lambda e / (S h c),$$

where F is the solar flux above the Earth's atmosphere, A is the effective area of the telescope, a is the area of a pixel (0.0625 arc sec²), and S is the area of the solar disk (2.88×10^6 arc sec²).

The efficiency of the telescope, filter, etc., is included in the efficiency factor e ; it is estimated to be 0.0645 (see section II.1.2). The solar spectral irradiance at 5250 Å in space is 2100 W/(m²μ), but this is reduced when observations are made in a solar absorption line. If we assume that most of the magnetic measurements will be made at 60 mÅ in the wing of the 5250 Å line, we can estimate the solar irradiance to be 1400 W/(m²μ). Then for $d\lambda = 0.120$ Å at $\lambda = 5250$ Å, the number of photons per second incident on a 0.25" pixel will be

$$N(\lambda) = 4 \times 10^6 \text{ photons per sec.}$$

Based on this parameter - the photons per pixel per second - the tradeoffs between various types of solid state sensors can be optimized to develop a camera system which will have the signal to noise and time resolution set by the scientific goals.

6.3 Description of Solid-State Sensors

When silicon absorbs a photon, an electron-hole pair is created. In solid state sensors, polysilicon electrodes are used to create potential wells in a silicon substrate so that photon produced electrons can be collected. There have been two main types of solid state sensors developed since the mid 1960s: charge-coupled devices (CCDs) and charge-injection devices (CIDs). There has been a great deal of research and development in the area of CCDs because of their use in computer memory products. Therefore, much of the discussion in this section will be on CCD sensors. From an imaging point of view, CID sensors have some significant advantages over CCD sensors. Unfortunately the electrode structure that provides the spatial integrity of the signal packet in a CID sensor also produces a higher amplifier noise level than that of CCD sensors which has been a severe impediment to their development.

6.3.1 CID sensors

The CID is an x-y addressable structure organized to read the signal charge through local charge transfer at each pixel. Injection into the underlying substrate is used to clear each pixel of signal charge and start a new signal integration period (monitor frame). The major advantage of a CID over a CCD array is due to this x-y addressable feature of the sensor, which allows immediate access and readout of the area of interest. In the CID sensor a single MOS gate separates the photosite from the video amplifier. In contrast, the pixels in the columns and rows preceding the area of interest in a CCD camera system have to be transferred through the video amplifier. Other advantages of the CID single gate readout structure are the absence of transfer loss effects and the isolation of pixel defects.

The major drawback to CID cameras is the increased noise level which is associated with the large capacitance (from the electrode structure) at the input of the video amplifier.

In the normal operation of the SAMEX magnetograph, x-y addressability is not an important feature since an active region must be confined within the detector array in order to carry out the extrapolation of the magnetic field into the corona (i.e., the boundary condition of a net zero magnetic field must be satisfied). The ability to address a smaller portion of the array is only important in studies requiring higher time resolution (e.g., studies of flare activity with changes in the magnetic field along a sheared neutral line).

In general the CID gives excellent signal isolation between individual pixels, has non-destructive data retrieval capability, can provide random access to groups of pixels, and has relatively simple topography.

Table 49 summarizes the advantages and disadvantages of the CID sensor.

Table 49. Summary of the advantages and disadvantages of the CID.

Advantages:

1. X-Y addressability
2. No transfer loss effects.
3. Defects are confined to each pixel.
4. Optical overloads are confined to each pixel.

Disadvantages:

1. Large output capacitance increasing noise level
(The output capacitance of CID sensors is in the range of 10-25 pF while the output capacitance for a CCD is typically in the range of 0.1-1.0 pF.)
-

6.3.2 CCD sensors

A CCD operates by transferring the signal charge generated at each pixel to the edge of the array where it can be processed. If the charge image is transferred across the photosensitive array while the scene is illuminating that array, smearing will occur. Three types of structural organizations of CCD sensors have been developed; some of these are more effective than others in reducing this smearing problem. The three types of structures are a staring sensor, frame-transfer sensor, and an interline-transfer sensor.

In a staring sensor structure, every pixel in the array is used to collect light. This type of sensor requires a shutter to eliminate image smearing during the readout of the image. In the SAMEX magnetograph, 775 images must be acquired in less than 300 seconds; using a shutter in such an operational sequence represents a potential failure point for a 3-year mission. Although several shutters could be used in the magnetograph design, the best approach would be to minimize the need for a shutter so that a shutter failure would not represent an instrument failure.

A frame-transfer sensor represents such a compromise. In this sensor the array is split into two equal areas: one for imaging and one for image storage. The storage area is covered by an opaque screen. This technique minimizes the need for a shutter since the image smearing is limited to the time required to move the exposed image from the image area to the storage area. The image shift time is a parallel transfer and is small when compared to the readout time which is a combination of parallel and serial transfers. Corrections to image smearing are limited to the parallel transfer direction, a factor which simplifies the numerical analysis required to correct the image. With this array, a shutter would be used as a calibration tool to correct for image smearing during the

parallel transfer; the shutter's duty cycle would thus be reduced and its failure potential minimized.

The interline-transfer sensor represents an approach to eliminate image smearing in CCDs. Each image row of the sensor has an adjacent storage row which is covered with an opaque material. With a single shift operation the entire image can be moved from the light collecting area to the readout area, effectively eliminating the need for a shutter. The disadvantage of this sensor is the loss in active collecting area which may be reduced to approximately 36% as compared to 89% for the staring or frame-transfer sensors.

The advantage of the CCD structure is the small capacitance which results in low amplifier noise. A disadvantage is the requirement that the charge transfer efficiency (CTE) must be very close to unity (>0.9999). This is especially important for very large arrays. To increase the CTE, a buried channel is used in many sensors to decrease surface trapping. Decreasing the surface trapping also allows the array to run at higher clock frequencies but reduces the signal handling capability of the device (Dereniak and Crowe, 1984). Another disadvantage of CCDs are defects. A local defect can affect the signal of every carrier packet that must be clocked through that area. A third disadvantage for standard CCD arrays is their sensitivity to optical overloads which cause electrons from one pixel to spill over into another (blooming). CCD arrays have been developed with a diode at every pixel that allows excess electrons to be drained off. However this reduces both the signal handling capability of the sensor and the light collecting area within the pixel.

The way the signal is clocked out of an array can affect the number of electrons a pixel can hold and the efficiency of the sensor in converting photons into electrons. Signal can be moved through a CCD array using either two-, three-, or four-phase clocking. Figure 54 shows a mechanical analog of the method by which the signal charge is moved through a CCD array using a three-phase clocking scheme. A four-phase clocking scheme is similar to three-phase clocking, but it has a 50% larger signal handling capability for the same size pixel (Barbe, 1975). Virtual phase CCDs, which are an extension of two phase CCDs, rely on the doping of the silicon substrate to force the signal carriers to move in the appropriate direction (Figure 55). In the virtual array, a dc bias gate is built into the surface of the silicon and a single polysilicon gate electrode can be used to move the signal charge through the array (for example, assume that gate 1 in Figure 55 is set at the dc bias and that gate 2 has voltage pulses that go below and above the gate 1 voltage). This simpler structure is very important in improving the quantum efficiency of front-illuminated devices while simplifying the clocking of the sensor (Hynecek, 1979). In Table 50 we have listed some of the performance advantages of the various CCD clocking structures.

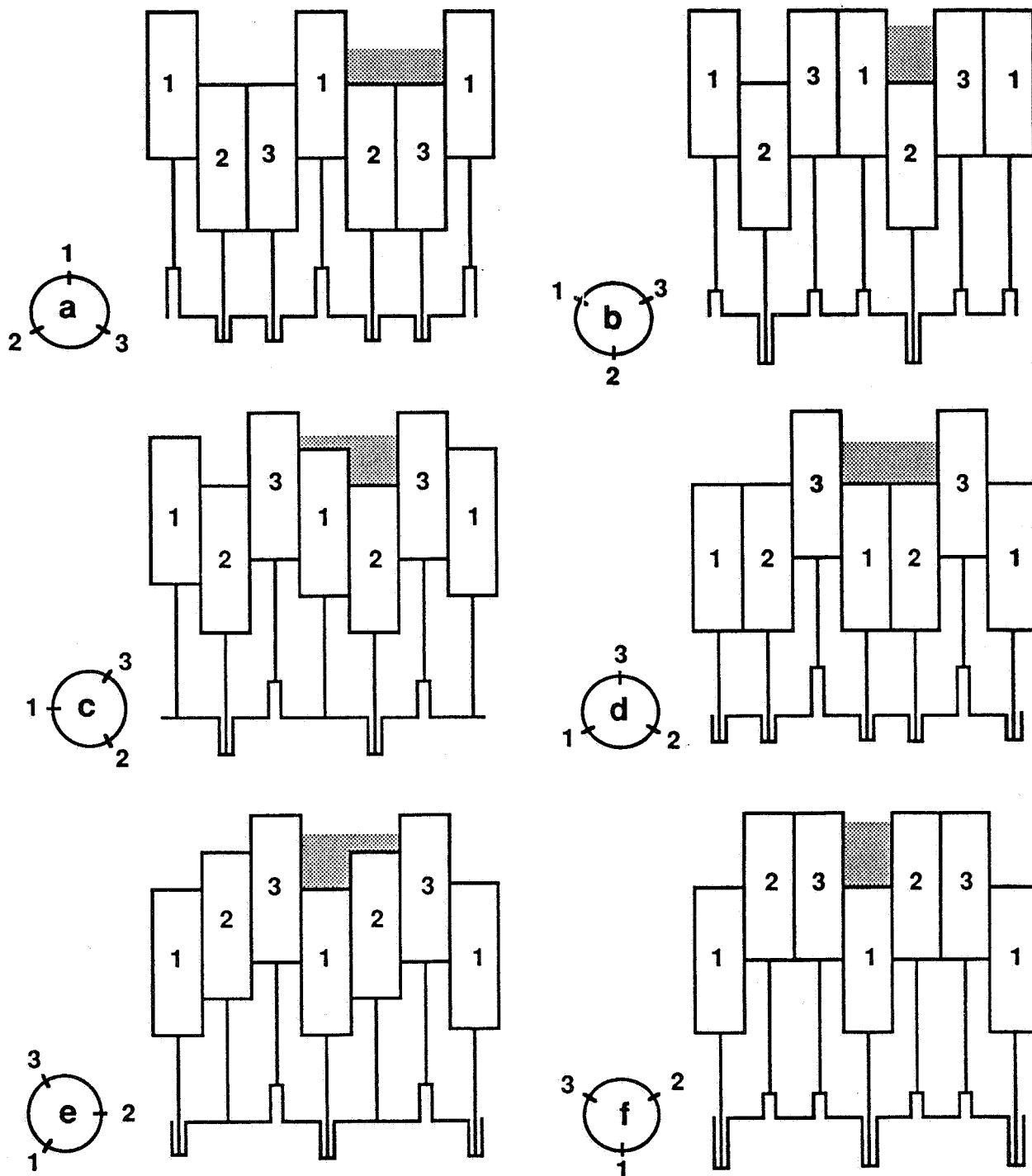


Figure 54. Three-phase clocking. This diagram shows a mechanical analog of three-phase clocking in which the signal is transferred from right to left. In a three-phase device, the direction that the signal packets move is determined by the timing of the three clocks.

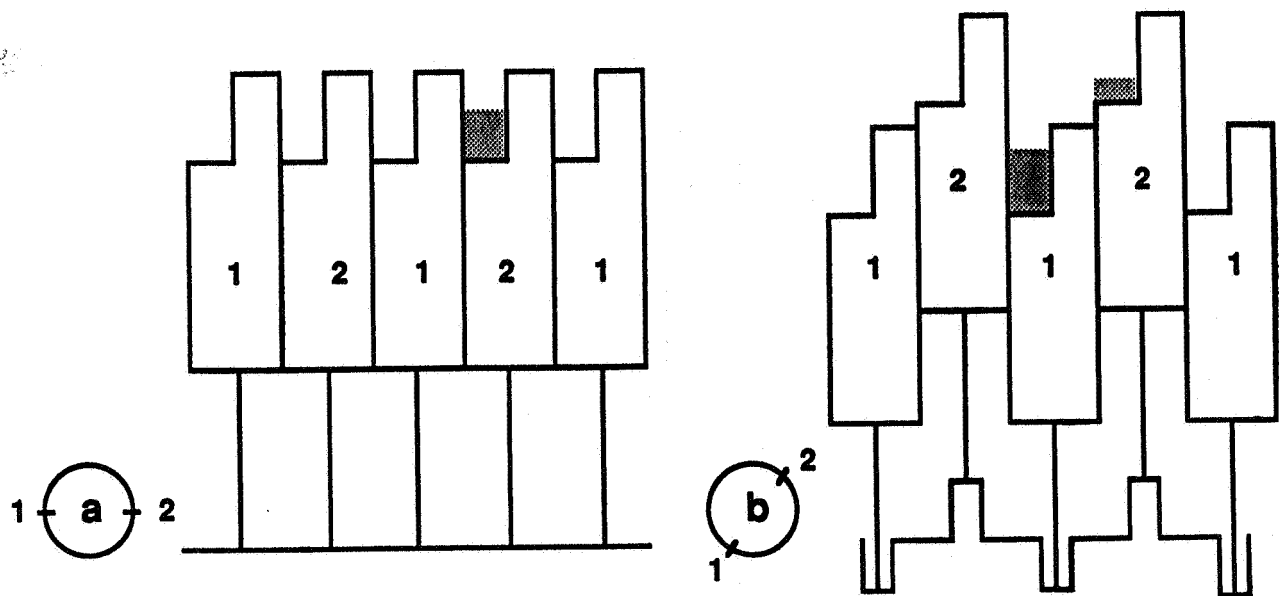


Figure 55. Two-phase clocking. This diagram shows a mechanical analog of two-phase clocking in which the signal is transferred from right to left. In this type device, the direction that the signal moves depends on the step function that is built into the pixels. This step function can be created by a diffusion implant in the substrate or by a variation of the oxide thickness between the electrode and substrate.

Table 50. Effects of different clocking structures on the performance of CCD arrays (Barbe, 1975).

CCD Structure	Performance Advantages
Two phase	Minimum number of transfers
Three phase	Highest packing density
Four phase	Largest signal handling capacity

6.4 Responsivity of Solid-State Sensors

Since the number of photons that are available for the detector system is fixed by the optical and scientific limitations discussed in sections 6.1 and 6.2, the exposure time and hence the time resolution of the instrument is determined by the quantum efficiency of the sensor (note: it is assumed that the readout time is smaller than the exposure time and that an image is read out as the next image is exposed). This section will discuss how different structures determine the "detectability" of the photons that are incident on the sensor. A photon is detected by a solid-state sensor when it is absorbed in the depletion region of a silicon substrate and creates one or more electron-hole pairs. The hole migrates out of the depletion region while the electron becomes trapped at the insulator interface. The quantum efficiency (η) of a detector is defined as the number of photons recorded by the detector divided by the incident photons and is a function of the incident light and the structural parameters of the photosite (Van de Wiele, 1976). In some of the literature, spectral responsivity is given instead of quantum efficiency. Spectral responsivity is the output signal (usually in amperes - current responsivity - R_i) divided by the radiant input (in watts). The relationship between the η and R_i is defined by the following equation (Dereniak and Crowe, 1984):

$$\eta = h c R_i / (q \lambda G) , \quad (6.1)$$

where λ is the wavelength in angstroms, h is Planck's constant, c is the speed of light, q is the electron charge, and G is the gain in the device and equals unity at optical wavelengths. The quantum efficiency of a solid-state sensor depends on several factors: the direction of the illumination with respect to the gate structure, the clocking structure of the device, and the charge transfer structure. In front-illuminated sensors, the light must travel through the polysilicon gate structure of the sensor which will absorb a portion of the light. This is very important for detectors that measure shorter (blue) wavelengths

where the diffusion length and the gate thickness determine how effectively the electrons are collected in the depletion region. CID sensors are similar to front-illuminated arrays in that light must travel through the electrode structure, but they have an advantage in that they do not need channel stops or pixel barriers (set by gate voltages or ion doping in the silicon substrate) to isolate the signal packets. This increases the light collecting area of the sensor. Back-illuminated sensors eliminate the absorption of light by the CCD electrode structure but are thin and usually require some sort of mechanical support.

6.4.1 Front-illuminated sensors

There is no difference in the quantum efficiency for staring and frame-transfer CCD arrays since the gate structure of the imaging area should be the same. The interline-transfer sensor has the lowest quantum efficiency because of the opaque strips in the image plane. The quantum efficiency for an interline-transfer sensor can range from 8% to 18% at the wavelength of 5250 Å (Table 51). For the frame-transfer (and staring) arrays, the quantum efficiency is a function of the clocking structure. The most efficient structure for a front-illuminated camera is the two-phase (or virtual phase) CCD sensor. By minimizing the polysilicon gate structure, this device allows more of the photons to reach the silicon substrate. Quantum efficiencies at 5250 Å can range from 18% for a four-phase device to 40% for a virtual phase. Of the front-illuminated devices only the virtual phase arrays have the sensitivity required to meet the scientific objectives, since the exposure time and hence the temporal resolution depend upon the quantum efficiency of the sensor. Other advantages that virtual arrays have over multiphase arrays are listed in Table 52.

CID sensors are similar to virtual phase front-illuminated arrays since light must travel through the electrode structure. Their relatively high quantum efficiency (Table 53) is due to the small polysilicon electrodes. As the pixel size decreases, the percentage of the cell covered by the polysilicon electrode increases; hence the decrease in the quantum efficiency.

Table 51. Quantum efficiencies QE of various CCD sensors.

CCD Sensor	Channel (a)	Clock (b)	Array Size	Pixel size (microns)	Illum. (c)	QE in % (5250 Å)
Amperex ¹	-	4	604 x 576	10.0 x 15.6	F	18
Fairchild ²	-	2	488 x 380	12.0 x 18.0	I	8
GEC (UK) ³	B	3	576 x 385	22.0 x 22.0	B-1	18
NEC ⁴	-	3	1280 x 980	9.9 x 9.8	I	13
Nippon ⁵	B	2	384 x 490	33.0 x 20.0	I	18
RCA ⁶	S	3	512 x 320	30.0 x 30.0	B	68
Tektronix ⁷	B	3	2048 x 2048	27.0 x 27.0	B-1	32
Tektronix ⁷	B	3	2048 x 2048	27.0 x 27.0	B-2	45
TI ⁸	B	3	800 x 800	15.0 x 15.0	B	50
TI ⁹	B	2	800 x 800	15.0 x 15.0	F	40
Videk ¹⁰	-	2	1320 x 1035	6.8 x 6.8	-	35

(a) Channel: B = buried, S = surface

(b) Clock: signal transfer is two- (2), three- (3), or four- (4) phase clocking

(c) Illumination: I = interline transfer, F = frame transfer, B = back-illuminated (either frame-transfer or staring sensors), B-1 = thick, B-2 = thin.

1. Sales information on the AMPEREX NXA1011 frame-transfer sensor

2. Sales information on the Fairchild CCD222 interline-transfer sensor

3. Wright and Mackay, 1981

4. Akiyama et al., 1986

5. Ishirara et al., 1980

6. Savoye, 1985

7. Blouke et al., 1985

8. Blouke et al., 1981

9. McGrath et al., 1983

10. Sales information on the Videk Megaplex sensor

Table 52. Comparison of virtual and multiphase arrays.

Advantages of Virtual Arrays:

- a. Single polysilicon gate provides a wider spectral response
- b. Technology of ion implantation is better understood than polysilicon gate technology.
- c. Since the polysilicon gate structure is simpler, smaller cells are possible.
- d. Minimizes the electronics required since only a single clock is required.

Advantages of Multiphase Arrays:

- a. The charge transfer direction can be bi-directional; this allows electrical signal to be stored in the array for calibration (could be useful when Sun is not visible).
- b. Multiphase arrays have a greater dynamic range than two-phase arrays (Vescelus and Antcliffe, 1976).

Table 53. Quantum efficiencies QE of various GE CID sensors.

Sensor	Array Size	Pixel Size (microns)	QE (5250 Å)
TN2700	512 x 388	13.6 x 23.4	48%
TN2710	512 x 776	13.8 x 12.0	30%

6.4.2 Back-illuminated sensors

A technique to improve the performance of the sensor is illumination of the backside of the array. Since the electrodes are on the front side of the silicon wafer, the electrode structure is not important in determining the quantum efficiency of the sensor. However, back-illumination is not compatible with the normal interline-transfer structure. With the elimination of the electrode structure, the primary limitation in the quantum efficiency of a back-illuminated array is the thickness of the array and the reflections from the sensor due to differences in the indices of refraction at the optical interface. Thin arrays improve blue sensitivity while thick arrays are used to detect red wavelengths. The quantum efficiency can vary from 20% to 70% at 5250 Å (Table 51).

In early back-illuminated CCDs, a window was cemented to the substrate of the sensor to increase their mechanical strength and to provide index matching which reduced losses due to

reflections. One problem associated with the structure is the production of interference fringes when exposed to monochromatic light. This problem increases with wavelength. At 5600 Å, a peak-to-peak variation of 16% can be seen in the signal output of a uniformly illuminated array. A window that has a 2° wedge has been shown to reduce the fringing to 3% (R. Aikens, Photometrics, Ltd., private communication). To eliminate interference fringes caused by thickness variations in the epoxy - CCD interface, the support structure (window) can be moved to the electrode side of the array (W. Hayward, Tektronix, Inc., private communication). This reduces the peak-to-peak fringing but lowers the quantum efficiency (index mismatch).

6.4.3 Section summary

Sensors that are back-illuminated have the highest quantum efficiency but may not be as "rugged" as other sensors and, depending on the physical construction of the supporting structure, and can have fringing problems when used with monochromatic light. CID arrays and virtual phase CCDs represent the strongest candidates for front-illuminated sensors. The leading candidate based on the scientific objective of a "high" time resolution vector magnetograph is the back-illuminated sensor with the frame-transfer structure.

6.5 Signal Transfer Characteristics of Solid-State Sensors

The resolution of a solid-state camera system is limited by several noise sources, by the charge mobility of the sensor, and by the amount of charge that can be stored in a pixel. The net effect is that most sensors have signal-to-noise ratios of $<10^3$. The scientific goals require a polarization sensitivity of 10^{-4} which means the digital images from the camera system must have a S/N of $>10^4$. If sensors had a S/N of 10^4 , then a single exposure would achieve the design goals of the instrument, and the time resolution would be determined by the exposure time (and readout time if the images were not processed in parallel) of the array. Unfortunately, available sensors have S/N values well below 10^4 , so that images of the same polarization exposure must be added together to improve the S/N of the measurement to the desired value. The number of images I (also referred to as enhancements) that must be added together to meet the 10^{-4} sensitivity is equal to:

$$I = (10^4 / (S/N))^2 ,$$

where S/N is the signal-to-noise ratio of the sensor. In Figure 56, the relationship of sensor signal-to-noise, image readout time, and number of enhancements required to meet the polarization and time resolution set by the scientific objectives of the instrument is shown.

In the following sections, we will discuss the major factors that determine the resolution of the camera system.

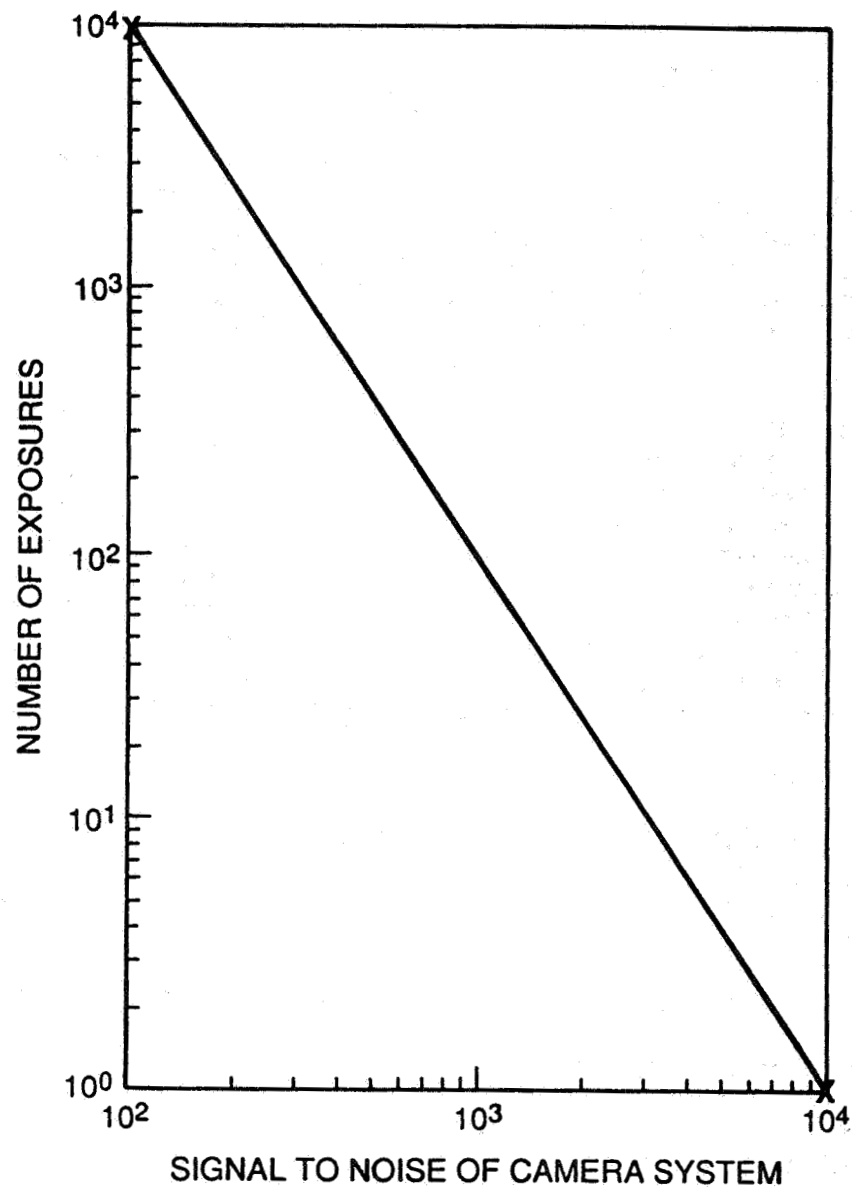


Figure 56. Image enhancements required for a given signal-to-noise ratio. This graph shows the number of image enhancements that would be required to achieve a polarization resolution of 10^{-4} for a detector system with a given signal-to-noise ratio.

6.5.1 Noise sources

Noise sources limit the dynamic range of the sensor which in turn determines the number of enhancements required, thereby affecting the time resolution. We classify noise sources for solid-state sensors in the following manner: background noise, transfer noise, and amplifier noise. In this section we discuss how the SAMEX detector system is affected by these noise sources and the tradeoffs that must be made to meet the design goals of the magnetograph.

There are several noise sources that are due to the manufacturing processes which will be neglected in this discussion since they can only be characterized through experimental measurements on individual sensors. These noise sources are associated with variations in the dark current (which will be discussed in the next section) across an array, local defects, and misalignment in the masks during fabrication. This type of noise is normally called fixed pattern noise. With proper calibration most of the fixed pattern noise can be cancelled to improve the S/N of the detector by as much as 15 db.

Background Noise

Background noise can be generated optically, electrically, or thermally. The optical noise is usually referred to as photon noise and is due to the random fluctuations in the number of photons incident on the sensor. This process is stochastic, governed by Poisson statistics, and the noise is equal to the square root of the number of photon-generated electrons. Although photon noise is shot noise in the true definition of the term, which is the random emission of light in time such that the arrival of photons follows a Poisson distribution, one often finds discussions in the literature about shot noise in terms of random motion of charges in electrical devices. The application of shot noise to solid-state devices is limited due to the fact that charge neutrality requires that, apart from the creation of electron-hole pairs, a new charge carrier can enter a device only as one leaves, so that the emission is not random. Therefore the term photon noise will be used to describe the optical shot noise of the sensor. Although photon noise can be the limiting factor on the signal-to-noise ratio of a sensor, much of the literature does not take it into account when discussing the resolution of the detector. Even though photon noise is not strictly a characteristic of the sensor it is the most important noise source (assuming all other noise sources have been minimized) for solid-state sensors operating near their saturation level (section 6.5.2) and must be included in any calculation of the overall S/N. The photon noise (which is normally given in terms of the electrons per pixel) is equal to

$$N = (2 G^2 n E_p A_d t)^{1/2}$$

where G is the photoconductive gain and equals 1 for solid-state sensors at optical wavelengths, η is the quantum efficiency, E_p is the photon irradiance, A_d is the pixel area, and t is integration time. The limitation that the photon noise places on the detector system arises from the fact that the well size, and hence the number of photons that can be recorded, for these devices is limited. Neglecting all other noise sources and assuming the photoconductive gain is unity at the wavelengths of interest, the maximum S/N from a single exposure of a solid-state image sensor would be

$$S/N_{\max} = WS/(WS)^{1/2}$$

where WS is the well size of the pixel assuming all electrons are photo-generated. Therefore, to obtain the 10^4 resolution in a single exposure would require a solid-state sensor with a pixel well capacity of greater than 10^8 electrons. This is approximately 2 orders of magnitude higher than the well capacity that is currently available with today's technology (Tables 54 and 55). Therefore the data acquisition system must add over 100 images to obtain 10^4 resolution using current solid-state detector technology.

Table 54. Comparison of electron capacities (Ne) for various solid-state sensors.

CCD Sensor	Channel (a)	Clock (b)	Array Size	Pixel Size (microns)	Ne	Ne per Micron ²
Amperex ¹	-	4	604 x 576	10.0 x 15.6	8.6×10^4	551
Fairchild ²	-	2	488 x 380	12.0 x 18.0	3.0×10^5	1389
GEC (UK) ³	B	3	576 x 385	22.0 x 22.0	5.0×10^5	1033
NEC ⁴	-	3	1280 x 980	9.9 x 9.8	6.8×10^4	701
Nippon ⁵	B	2	384 x 490	33.0 x 20.0	4.6×10^5	697
RCA ⁶	S	3	512 x 320	30.0 x 30.0	3.5×10^5	389
Tektronix ⁷	B	3	2048 x 2048	27.0 x 27.0	7.0×10^5	960
TI ⁶	S	3	400 x 400	25.0 x 25.0	3.0×10^5	480
TI ⁶	B	3	800 x 800	15.0 x 15.0	7.0×10^4	311
TI ⁸	B	2	800 x 800	15.0 x 15.0	1.3×10^5	577
TI ⁹	-	-	1024 x 1024	18.0 x 18.0	2.0×10^5	617
Videk ¹⁰	-	2	1320 x 1035	6.8 x 6.8	7.0×10^4	1513

(a) Channel: B = buried, S = surface

(b) Clock: signal transfer is two- (2), three- (3), or four- (4) phase clocking

1. Sales information on the AMPEREX NXA1011 frame-transfer sensor
2. Sales information on the Fairchild CCD222 interline-transfer sensor
3. Wright and Mackay, 1981
4. Akiyama et al., 1986
5. Ishirara et al., 1980
6. Gunn and Westphal, 1981
7. Blouke et al., 1985
8. McGrath et al., 1983
9. Communication with R. Aikens, Photometrics, Ltd.
10. Sales information on the Videk Megaplug sensor

Table 55. Saturation signal and amplifier noise for a CID 320 x 244 array as a function of pixel size (average density for CID cameras is 1384 electrons/micron²).

Average Pixel Size (microns)	Noise Level (electrons)	Saturation Level (electrons)
34	1.1×10^3	1.6×10^6
22	5.4×10^2	7.0×10^5
18	4.3×10^2	4.2×10^5
13	3.2×10^2	2.5×10^5
10	2.1×10^2	1.1×10^5

Values are derived from chip area in Figure 6 of Brown et al. (1980) assuming square pixels. The temperature was not given; therefore, noise was assumed to be dark current at room temperature.

The electrically generated noise (or input noise) is associated with the input gate signal that is used to create a "fat zero" to eliminate any empty interface states or bulk traps and thus minimize transfer noise. This type of noise is similar to the amplifier reset noise. Input noise is more important for sensors that are exposed to low light levels. Due to the large signal levels expected from observations with the SAMEX magnetograph (except during limb scans), a "fat zero" signal should not be required and this type of noise will be neglected (the input gate noise should be included for transfer efficiency calibrations which use it as an input signal source).

Thermally generated noise is usually referred to as dark current and is associated with carriers which are thermally generated that bring the potential well into equilibrium. Uniform dark current can be cancelled digitally (or electrically) by subtracting a background image from the exposed image. Dark current varies with temperature and can be expressed in the following form (Blouke et al., 1981):

$$J_{dc} = C T^m \exp(-E_g/n k T), \quad (6.2)$$

where J_{dc} is the dark current (nA/cm²), C is a constant, T is temperature (°K), k is the Boltzmann constant (8.62×10^{-5} eV/°K), m is 3/2 for silicon, n is 2 for silicon, and E_g is the silicon band gap energy and is given by $E_g = 1.1557 - (7.021 \times 10^{-4} T^2)/(1108 + T)$. The number of thermally generated electrons per pixel, S_{dc} , is obtained from the dark current J_{dc} from the relation

$$S_{dc} = J_{dc} A_d/q \quad (6.3)$$

where A_d is the pixel area and q is the electron charge (1.6×10^{-19} Coulombs). Since the thermal noise is stochastic and described by Poisson statistics, it is equal to

$$N_{dc} = (S_{dc})^{1/2}.$$

Typical dark current densities J_{dc} for three- and four-phase CCD arrays have been quoted to be 10 nA/cm^2 at 25°C (sales information on Tektronix CCD imagers; Wight, 1975). With this value, the constant C in Equation (6.2) can be determined, and the number of thermally generated electrons can be plotted as a function of temperature (Figure 57). Near room temperature, the dark current will double with a 10°C change in temperature (Jorden and van Breda, 1981; Fairchild CCD222 sales literature; Videk sales literature).

Although 10 nA/cm^2 is the typical value for dark current in multiphase CCDs, CID sensors and virtual phase CCDs have lower values. The dark current associated with CID cameras, 2.5 nA/cm^2 , is approximately 4 times lower than that for multiphase CCDs (Aikens et al., 1976). The dark currents are lower because the thermal generation of charge in the undepleted region of bulk silicon (photon charge generation area) is less than the charge generated in the depleted storage area. In the CID sensor, the photon charge generation area is larger than the charge storage area which accounts for the lower dark current. Separate storage areas in CCD sensors (frame and interline transfer) also add to the dark current (Burke and Michon, 1976).

The development of virtual phase arrays has been accompanied by a further decrease in dark current. Values ranging from 0.58 nA/cm^2 to as low as 0.01 nA/cm^2 have been reported (Luppino et al., 1986). This reduction is brought about by filling the interface states in the virtual region with holes and the creation of an inversion layer which will provide holes to fill any surface traps when the gate is biased to a large negative voltage, ~ 10 volts.

Dark current in multiphase CCDs can be reduced to levels comparable to virtual arrays by reducing the clocking voltages (R. Aikens, Photometrics, private communication). Since this will reduce the well capacity significantly (for example a 7×10^5 electron well will be reduced to 3×10^4 electrons), this approach is best suited for simple photon counting.

Dark current has two effects on the signal-to-noise calculations. It decreases the available well capacity for photogenerated electrons while acting as a shot noise source. The data in Figure 58 show how the S/N will vary with temperature, assuming a full well of 7×10^5 electrons. In this case we are showing only the optical and dark current noise sources. The signal-to-noise can be written:

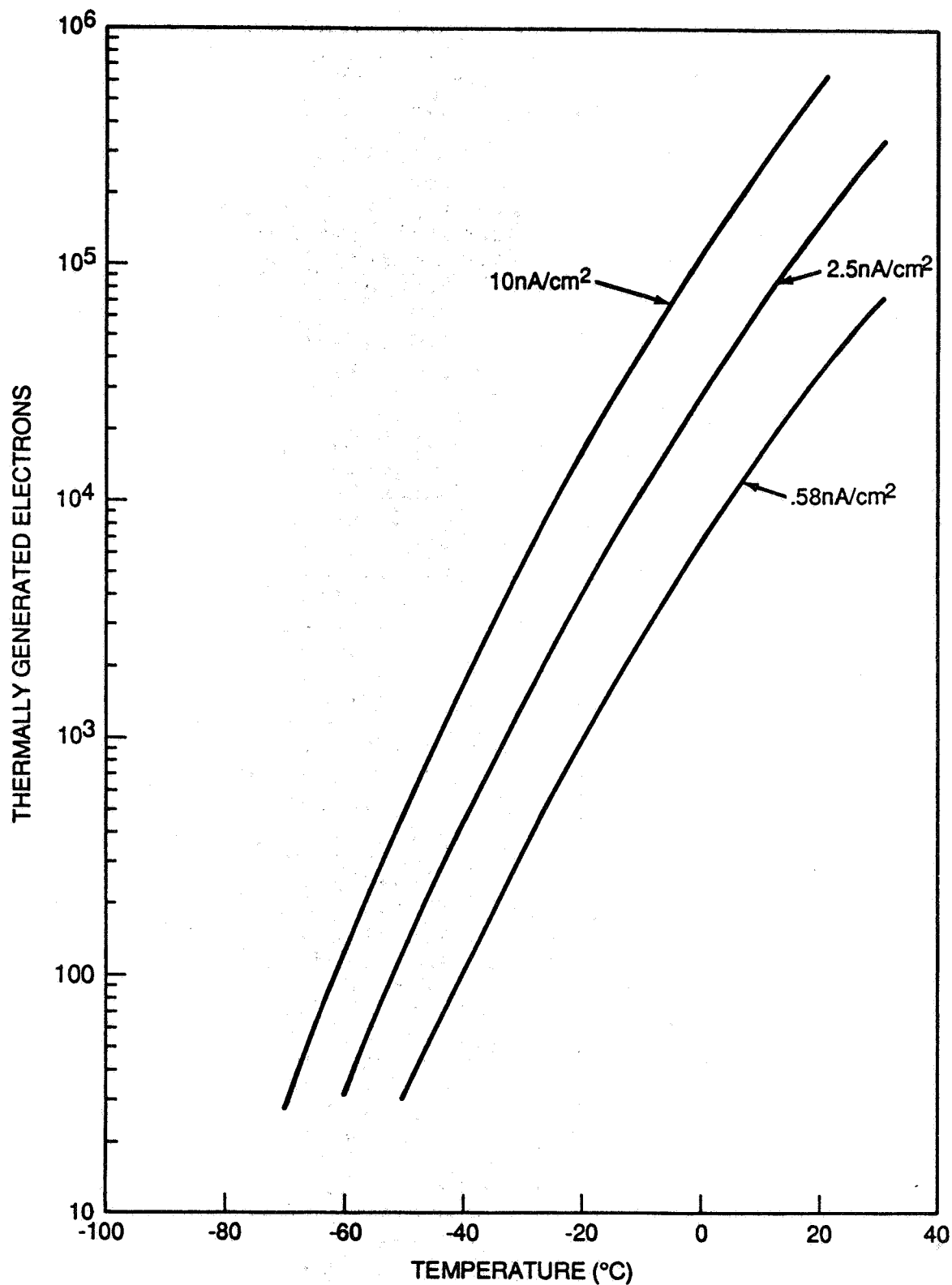


Figure 57. Thermal electron generation. These curves show the relationship between thermally generated electrons and temperature for three different dark current densities. A pixel size of 27 x 27 micron² and a readout time of 300 milliseconds were assumed.

$$S/N = (WS - S_{dc}) / (N_{dc}^2 + N_p^2)^{1/2}.$$

Since it is assumed that the well is filled with either photogenerated electrons or dark current, the total noise is equal to the square root of the well size (WS). Therefore, the dark current begins to become important to the signal-to-noise calculations (Figure 58) when it is larger than 1% of the well capacity. Although virtual phase arrays have the lowest dark current density, cooling the array to -20 °C would minimize that advantage at the expense of providing a cooling system.

Transfer Noise

The transfer noise in CCDs can be generated by charge loss, charge trapping, and by gate clocking. A great deal of research and development in CCD sensors has been directed toward reducing transfer noise since it can determine the dynamic range of the sensor. Although an incomplete transfer of the signal packet in a CID sensor is not as critical as the transfer of signal between pixels in a CCD, this problem can affect the polarization resolution of the magnetograph. If residual signal remains on the sensor between polarization measurements, this signal will represent a source of cross talk in the measurements. Therefore, the signal stored in the CID sensor would have to be erased before data at the next polarimeter position are acquired.

The transfer noise of a CCD sensor depends upon the charge mobility which will be discussed in the next section. This noise source is determined by the charge transfer efficiency which is related to (1) the silicon-silicon dioxide interface states, (2) the bulk traps in the substrate, (3) the thermal diffusion rate for small signal packets, (4) the self-induced drift (electrostatic repulsion) for large signal packets, and (5) the electric field on adjacent gates. The last three factors determine the maximum clocking that a sensor can have for a given charge transfer efficiency; the first two determine the actual fluctuation in the signal charge assuming a "complete" transfer. The noise associated with the transfer loss, N_{t1} , is given by

$$N_{t1} = [2 e N_g (N_s + N_b)]^{1/2},$$

where e is the fractional loss per gate, N_g is the number of gates, N_s is the number of signal carriers (optical and dark current), and N_b is the number of background carriers (at zero). The fractional loss per gate (e) must be independent of the amount of charge to be transferred (Carnes and Kosonocky, 1972). This will be discussed in greater detail in the section on charge mobility (section 6.5.3).

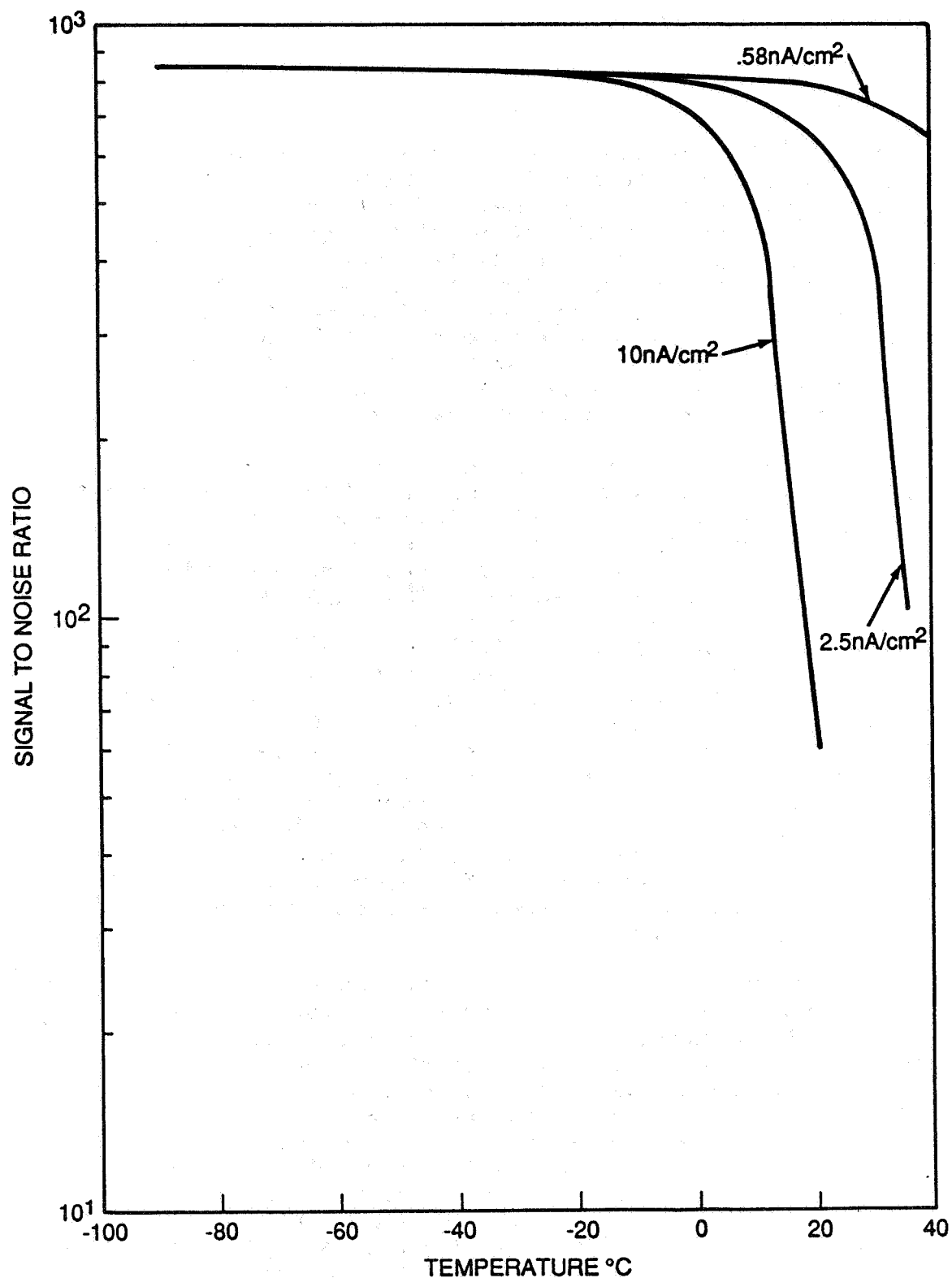


Figure 58. Dependence of signal-to-noise ratio on temperature. These curves indicate how the signal-to-noise ratio of a detector system varies with temperature for three typical dark current densities. A saturation level of 700×10^3 electrons and a readout time of 300 milliseconds were assumed.

Amplifier Noise

The third noise source to be discussed is associated with the output amplifier. This noise includes the reset noise from the gates that transfer the signal and reference voltage to the amplifier and the on-chip MOS field effect transistor (MOSFET). The first stage amplifier is normally located on the sensor to reduce the load capacitance. One of the major problems of CID sensors has been the high load capacitance which produces a large kTC noise source. Recent developments in CID cameras have minimized the kTC noise source by changing the readout procedure (Hynecek, 1979). This in turn has increased the S/N to nearly 50 dB which is comparable to many CCD sensors.

Amplifiers for solid-state sensors can have destructive or non-destructive readouts of the signal. There are two types of amplifiers in the destructive classification: current sensing and floating diffusion. In the current-sensing amplifier, the output gate is connected to a resistor or a transimpedance amplifier which changes current into voltage. This type is linear and low in noise but limited to low frequency because of high stray capacitances. In the floating diffusion amplifier (the most commonly used), the voltage applied to the output amplifier is determined by selection of a reference gate or signal gate. The reference gate controls the application of the reference voltage to the output amplifier while the signal gate controls the clocking of the data through the serial register to the output amplifier. By multiplexing the two gates, the reference voltage is applied to the output amplifier and a sample taken; then the signal packet is added to the output and a second sample is made. The difference between the two samples represents the true signal. This technique is called correlated double sampling and is used to minimize noise.

The two most common types of non-destructive amplifiers are the floating gate and the distributed floating gate amplifiers. Non-destructive amplifiers have a sensing gate between the normal electrode structure and the silicon substrate that will sense charge packets as the signal ($V = Q/C$) is clocked past it; this signal is then compared to a reference source. Since the reference voltage is used mainly to replace any charge lost by leakage in the MOS amplifier, the reset gate clocking frequency can be adjusted so that it is not in the bandpass of the preamp. This will eliminate the reset gate as a noise source. The distributed floating gate amplifier is the same as the floating gate, but since the charge is not destroyed, it is sampled at several points. This approach is useful in surface channel CCDs (Carnes and Kosonocky, 1972) but does not have the sensitivity in buried channel devices where the signal packets are deep in the silicon substrate. Since most CCDs utilize a buried channel to increase the charge transfer efficiency, the floating gate amplifier, which is more difficult to manufacture, has become less important. One possible solution is to combine a buried channel array with a surface channel amplifier.

The amplifier noise is determined by the transconductance of the MOSFET and the circuitry required to move the signal from the shift register to the input of the amplifier (Dereniak and Crowe, 1984). If amplifiers were ranked according to their output noise level, the floating diffusion amplifier would have the highest noise level followed by the floating gate and distributed floating gate amplifiers.

The choice of an amplifier also depends on the dynamic range of the sensor and the sensitivity to the signal packets. Generally, low noise implies a reduced amplifier saturation threshold, which reduces the dynamic range (Wight, 1975). A reduced dynamic range implies more image enhancements would be required to obtain a sensitivity of 10^{-4} in the polarization measurements. Since photon (shot) noise will be the dominant noise source, the floating diffusion amplifier has the sensitivity and dynamic range required to handle the large signal levels that are needed to minimize this noise source. The noise N_a associated with this type of amplifier is given by

$$N_a = (N_r^2 + N_m^2)^{1/2}.$$

In this expression, N_r is the noise introduced by the resetting of the MOS channel to the reference voltage and N_m is the MOSFET preamplifier noise. N_r and N_m are determined from the following equations:

$$N_r = (k T C_o / q^2)^{1/2},$$

and

$$N_m = [(C_o^2 8 k T B) / (3 q^2 g_m)]^{1/2}.$$

In these equations, C_o is the output capacitance, B is the bandwidth ($1/2$ the clocking frequency - Nyquist rate), and g_m is the transconductance ($1000 \mu\text{mho}$). For CCDs, the output capacitance is between 0.1 and 1.0 picofarads. One of the limitations to early CID cameras was the high output capacitance which was typically 25 pF (Aikens et al., 1976). Advances in the late 1970s have reduced this output capacitance, making CID sensors more competitive than their early counterparts (Brown et al., 1980). The data in Table 56 show how capacitance and temperature affect noise in a floating diffusion amplifier. The floating gate amplifier eliminates the reset noise but, in a buried channel CCD, suffers a loss in sensitivity due to the fact that the signal charge is "deep" in the substrate. Since correlated double sampling is a technique that can help suppress the reset noise in the floating diffusion amplifier and since there is no loss in the readout sensitivity, it is the amplifier that is normally used in buried channel CCDs (BCCD).

Table 56. Effects of temperature and capacitance on the floating diffusion amplifier noise (the noise is rms carrier fluctuation).

Temperature (°K)	N_r	N_m	N_a
Capacitance = 0.1 pF			
216	104.8	3.8	104.8
236	109.5	4.0	109.5
256	114.1	4.2	114.1
276	118.5	4.3	118.5
296	122.7	4.5	122.7
Capacitance = 25 pF			
216	1657	957	1914
236	1732	1000	2000
256	1803	1041	2082
276	1873	1081	2162
296	1940	1120	2240

For a floating diffusion amplifier with correlated double sampling or a floating gate amplifier, the reset noise N_r could be neglected.

Summary

If the background, transfer and amplifier noise sources are independent, then the total noise can be represented by the square root of the sum of the squares of all the noise sources (Dereniak and Crowe, 1984):

$$N_{\text{tot}} = (N_r^2 + N_m^2 + N_{\text{tl}}^2 + N_{\text{dc}}^2 + N_p^2)^{1/2}.$$

In Figures 59-62, the signal-to-noise characteristics of three different CCDs and a CID are illustrated. These figures are discussed more fully in section 6.6.3.

6.5.2 Well size

The relationship between well capacity and pixel size is set by the dielectric strength of the silicon dioxide gate insulator. For a surface channel sensor, the number of electrons per unit area, N_{max} , that can be stored on a MOS capacitor is given by (Barbe, 1975):

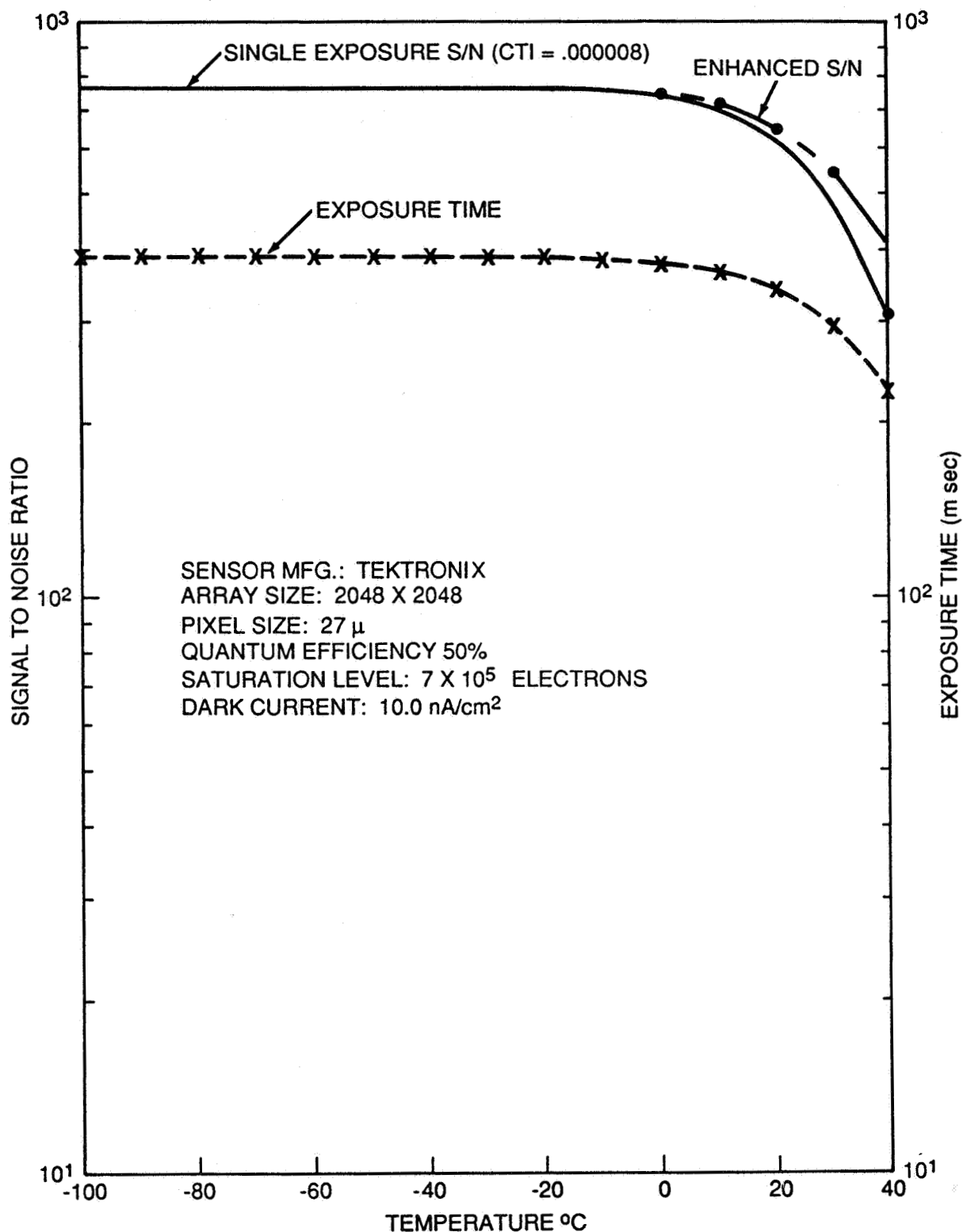


Figure 59. Signal-to-noise characteristics of a Tektronix CCD. The CCD is a back-illuminated, 1024 x 1024 frame-transfer device. Parameters assumed in the calculations for these curves include a pixel size of 27 x 27 micron², a well size of 700 x 10³ electrons, a quantum efficiency of 50%, a dark current of 10 nA/cm², a capacitance load of 0.1 pF at the output amplifier, and a charge transfer inefficiency (CTI) of 5 x 10⁻⁵.

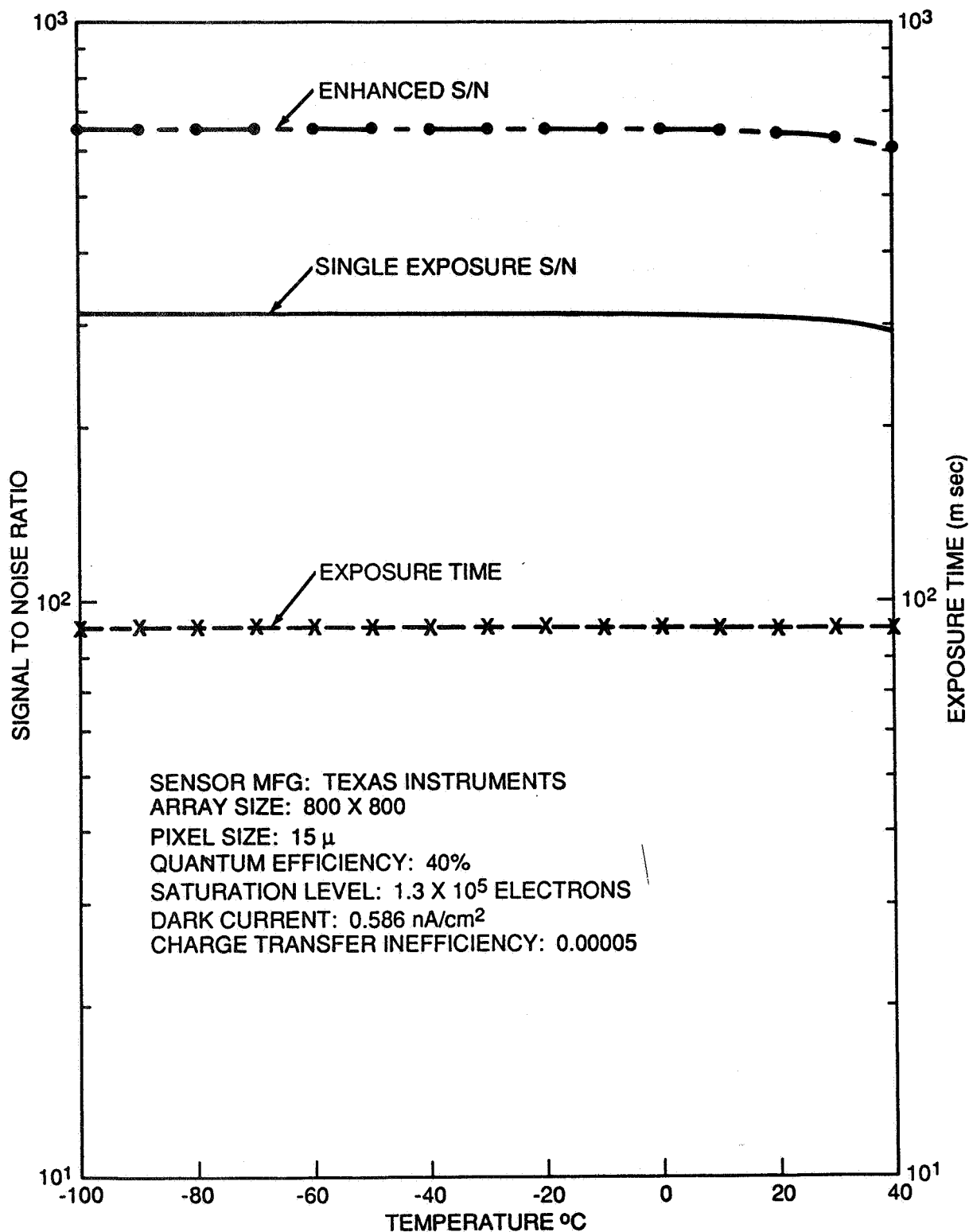


Figure 60. Signal-to-noise characteristics of a Texas Instrument CCD. The CCD is a front-illuminated, 800 x 800 (assumed frame transfer) device. Parameters assumed in the calculations include a pixel size of 15 x 15 micron², a well size of 130×10^3 electrons, a quantum efficiency of 40%, a dark current of 0.6 nA/cm², a capacitance load of 0.1 pF at the output amplifier, and a charge transfer inefficiency (CTI) of 5×10^{-5} .

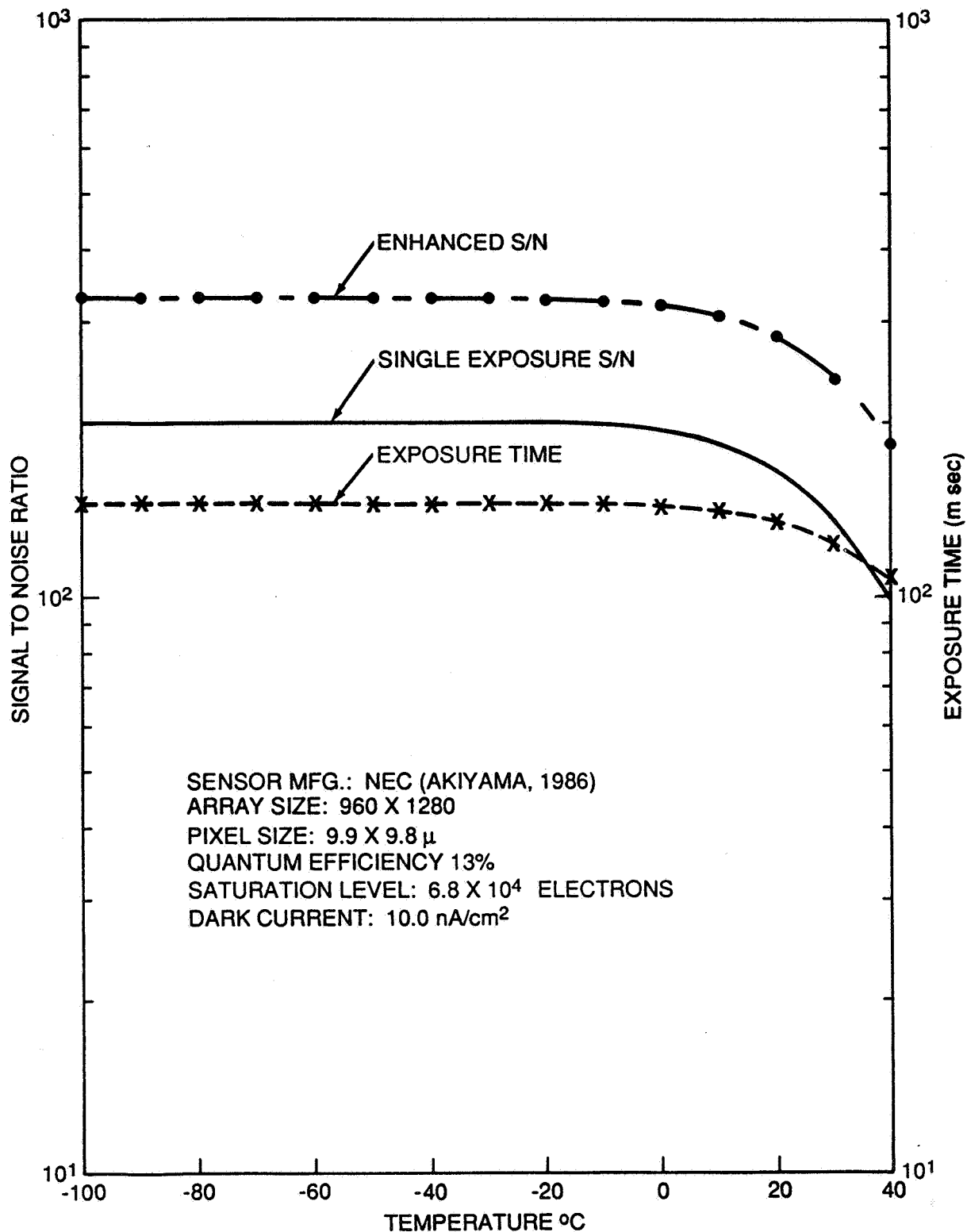


Figure 61. Signal-to-noise characteristics of a NEC CCD. The CCD is a front-illuminated, 960 x 1280 interline transfer device. Parameters assumed in the calculations include a pixel size of 9.9 x 9.9 micron², a well size of 68×10^3 electrons, a quantum efficiency of 13%, a dark current of 10 nA/cm², a capacitance load of 0.1 pF at the output amplifier, and a charge transfer inefficiency (CTI) of 5×10^{-5} .

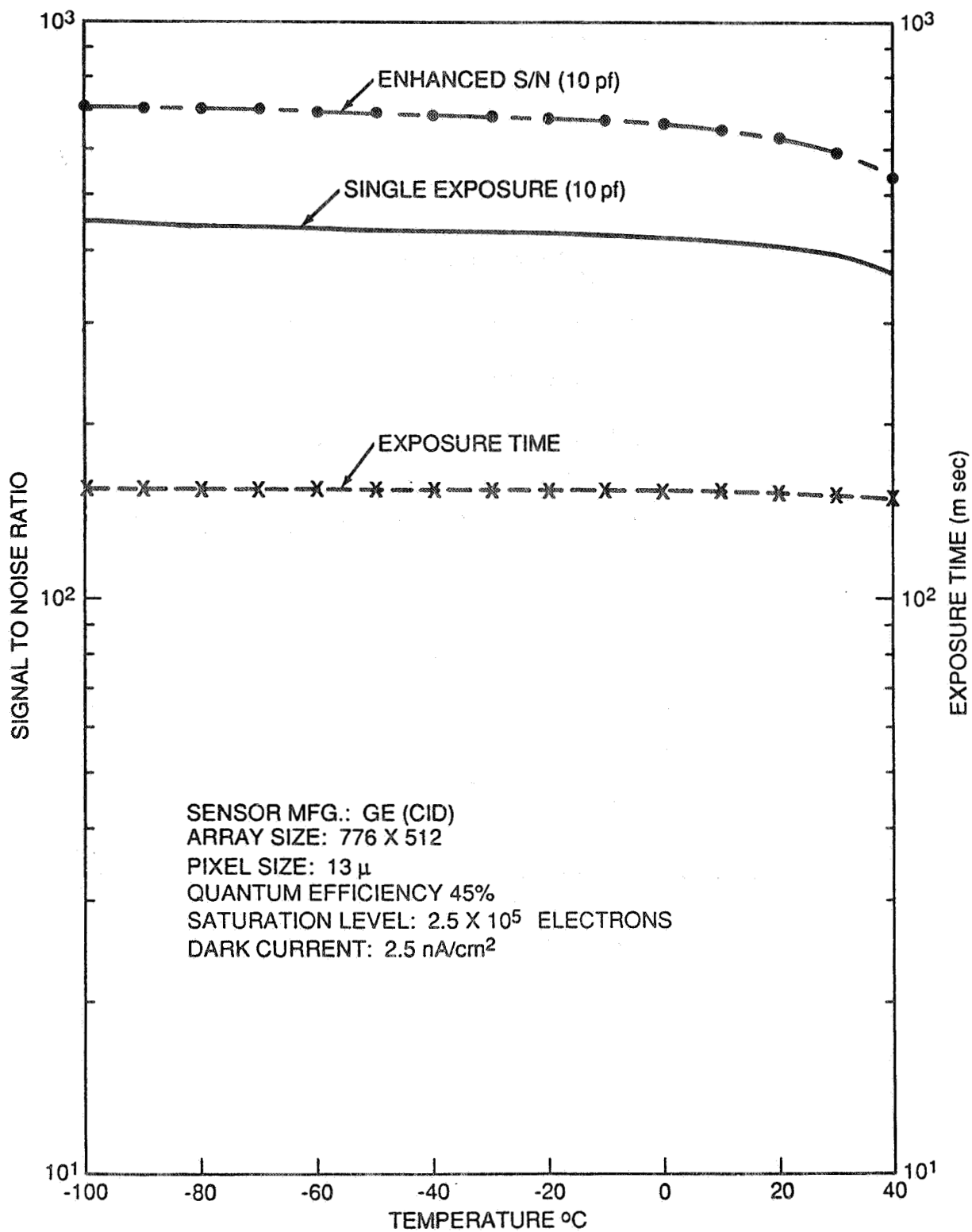


Figure 62. Signal-to-noise characteristics of a General Electric CID. The CID is a front-illuminated, 776 x 512 device. Parameters assumed in the calculations include a pixel size of 13 x 13 micron², a well size of 250×10^3 electrons, a quantum efficiency of 45%, a dark current of 2.5 nA/cm², a capacitance load of 10 pF at the output amplifier, and a charge transfer inefficiency (CTI) of 0.2.

$$N_{\max} \approx (D_{\text{ox}}/d) V_g/q,$$

where D_{ox} is the dielectric strength of the insulator, d is the insulator thickness, and V_g is the gate potential. Assuming that the insulator thickness is 1000 Å and the gate voltage is 4 volts, the electron density for a surface channel sensor would be 8×10^{11} electrons/cm² or 8×10^3 electrons/micron².

The electron density that a buried channel sensor can accommodate is a complicated function of the thickness of the channel, the doping density of the channel, the depletion voltage applied, and the gate voltage. A derivation of the equation for the electron density for a buried channel device can be found in Sequin and Tompsett (1975). For a 3-micron channel whose doping density is 3×10^{16} electrons/cm², the number of electrons that the device could hold is 6×10^{11} electrons/cm² or 6×10^3 electrons/micron² (buried channel electron density).

In imaging arrays, the electron density that a cell will hold will be reduced even further by both the channel stops between columns and by the barrier gates required to isolate the individual pixels. In Table 54 we list the pixel size, well capacity, and the electron density for several CCD sensors. The average density listed in Table 54 is 726 electrons/micron², which is approximately 10% of the calculated density. The well capacity of CID sensors as a function of pixel size is listed in Table 55. The average electron density for a CID sensor is 1384 electrons/micron². Large electron densities are important in order to increase the signal-to-noise ratio of the sensor and thus reduce the number of image integrations required to obtain the 10^{-4} polarization resolution. CID camera systems have been reported with well capacities as high as 6×10^6 electrons (Aikens et al., 1976). Electrostatic repulsion can be a problem when filling a well to its maximum capacity. A non-linear response has been noted in CID cameras after the well reached 50% of its saturation level of 2×10^6 (Jorden and van Breda, 1981).

The variation in the electron density of CCD sensors (Table 54) is probably due to differences in the definition of linear response. For example when there are 3×10^5 electrons in the Tektronix 2048 x 2048, the deviation from a linear response is 0.5% (electrons read out versus photons absorbed) and at 7×10^5 electrons the deviation is 3.0% (Brian Corrie, Tektronix, private communication). While the channel stops and isolation gate voltages set the saturation level (the point where signal spills over into adjacent pixels), the electrostatic repulsion and cell size determine the linear response. In our application weak magnetic fields have high signal levels (photospheric fields) and a non-linear response will reduce the sensitivity to those magnetic structures.

6.5.3 Charge mobility

Any discussion of charge mobility in a CCD must include its effect on transfer inefficiency. For a surface channel CCD (SCCD), the fundamental limitation is set by the decay time constant which is determined by the thermal diffusion of the signal packet. For large signal levels, the self-induced fields from the electrostatic repulsion of charges will dominate the transfer. This dominance will continue until the self-induced field produced by the electron density equals the thermal field. At this point the final transfer will be governed by the decay constant of thermal diffusion (Sequin and Tompsett, 1975).

The self-induced fields are the longitudinal electric fields produced by the charges themselves. These fields can help or inhibit the transfer of charge, depending on their position in the depletion region. In Figure 63 (Kosonocky and Carnes, 1973), the variation of charge transfer inefficiency as a function of clock frequency is shown for a SCCD. Although a sensor using a "fat zero" will reduce the signal capacity of a pixel, it can decrease the charge transfer inefficiency. Buried channel devices have an effect similar to sensors using a fat zero. The doping material in the buried channel device decreases the surface and bulk traps so that the charge transfer inefficiency can be reduced which in turn reduces the signal capacity of a pixel. Therefore at clock frequencies below 1 MHz the charge transfer inefficiency is dependent on how the sensor handles the surface and bulk traps while at higher frequencies the thermal diffusion time constant determines the charge transfer inefficiency. The exponential decay of a charge packet (assuming no fringe fields, to be discussed later in this section) in a pixel is determined by the thermal diffusion time constant described by the following relation (Sequin and Tompsett, 1975):

$$\tau = 4 L^2 / \pi^2 D ,$$

where L is the electrode length of the gate and D is the diffusivity of the carriers. The electron diffusivity can be determined using the Einstein relationship (Bar-lev, 1979)

$$D = k T \mu / q$$

where D is the electron diffusivity and μ is electron mobility ($0.135 \text{ m}^2/\text{V-sec}$ @ 300°K). Therefore, the electron diffusivity at 300°K is $3.3 \times 10^{-3} \text{ m}^2/\text{sec}$. For a cell whose length L is 23 microns, the time constant would be 64 nsec. As expected, the inefficiency starts to rise sharply (Figure 63) as the transfer time is reduced to the order of the decay time constant.

Although the thermal diffusion time constant governs the charge transfer in square potential wells, fringing fields from adjacent electrodes produce electrical fields with longitudinal components. If properly designed these fields can decrease the transit time significantly. It is the strong fringing fields

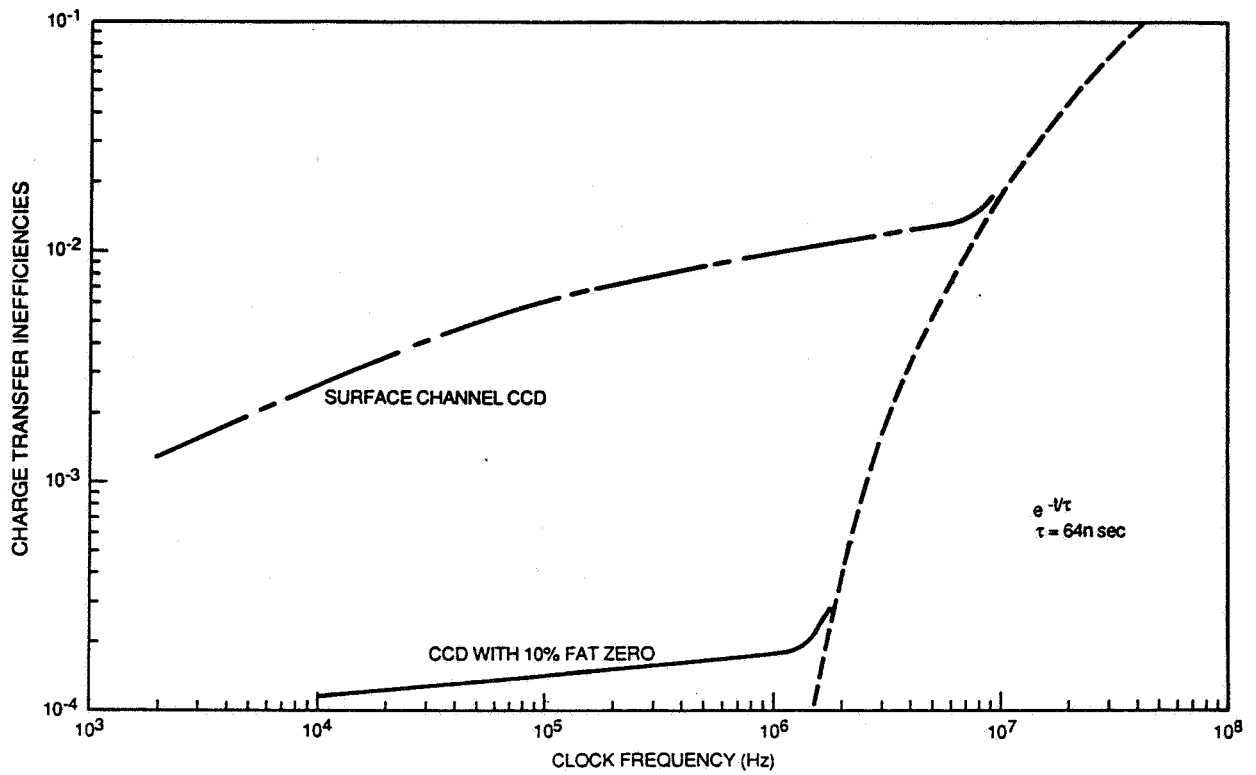


Figure 63. Dependence of charge transfer inefficiency on clock frequency. These curves show the relationship between the clocking frequency and charge transfer inefficiency for surface channel CCDs with fat zero signals of 0% and 10%.

that give CCDs their superior charge transfer efficiencies at high frequencies. For buried channel CCDs which use a doping material with a polarity opposite that of the substrate (this reduces the interface trapping and produces signal packets that are made up of majority carriers; SCCD have signal packets that are minority carriers), the doping level must be optimized to reduce interface traps while minimizing the smoothing effect it has on the potential profile of the cell. Longitudinal fields can reduce the transit time (time required to move the signal between two pixels) to less than 10 nsec (Sequin and Tompsett, 1975) which is an order of magnitude faster than CCDs that do not utilize fringing fields (~100 nsec for the same charge transfer inefficiency of 10^{-4}).

For large arrays, charge mobility (which includes thermal diffusion, fringing fields, and charge-induced drift), while providing a theoretical limit for the transfer time between gates, becomes important only in very large pixels (>52 microns for a four-phase device; Janesick, JPL, private communication). Most CCDs fail to operate at their theoretical limit because of the power dissipated in the clock drivers, insufficient bandwidth of the on-chip amplifier (Deyhimy et al., 1979) and, for large arrays, transmission line effects in the parallel transfer polysilicon gates.

6.5.4 Transmission line effects on readout times

For large arrays the limiting factor on the readout time is the time delay on long polysilicon gates which behave like RC transmission lines. Assuming a parallel gate is 2.5 mm long by 10 micron wide and has a sheet resistance of 80 ohms per square (the term "square" is a pure number having no dimensions and is related to the number of squares of width w that will fit in a thin film resistor of length l and width w ; Berry et al., 1968), the delay between the voltage applied at the beginning and the end of the gate across the channels can vary from 70 to 120 nsec, depending on the magnitude of the clock voltages and gate resistance (Kosonocky and Carnes, 1973). One of the arrays that is being considered for the SAMEX detector system is the Tektronix 2048 x 2048 ($55 \times 55 \text{ mm}^2$) which has parallel transfer gates 25 times as long as the array studied by Kosonocky and Carnes. The high resistance of the polysilicon gates will cause both a delay in the clock pulse as it propagates through the array and a decrease in the gate potential. Using existing polysilicon technology, the parallel shift time for a standard 512×512 array (23 micron pixels, 12 mm long parallel gates) can vary from 1 to 20 microseconds depending on the charge transfer efficiency needed (Allen Delamere, Ball Brothers, private communication). For a 2048 parallel gate structure with 27 micron pixels, the transmission line is 4 times longer; therefore, the parallel shift time could be as long as 80 microseconds for high charge transfer efficiencies.

Therefore, increasing the conductivity of the polysilicon gate structure is one of the important research areas in the development of large CCD arrays. One way to reduce the transmission line effects for a back-illuminated CCD is to use metal electrodes across the sensor instead of polysilicon. Although metal electrodes have been used in CCD sensors, the process has not been developed for large arrays and would be a departure from the standard manufacturing techniques. Another possible technique is the creation of metal busses along the parallel gate structure. The number of polysilicon gates between the metal leads would determine the parallel shift time. This has the disadvantage of adding another layer of complexity to the manufacturing process, but has the advantage of using the "standard" structure.

6.5.5 Location of Amplifier

As discussed earlier, a "buried channel" structure is used to reduce surface trapping in the silicon. Unfortunately this also increases the gate capacitance and hence the power dissipated in the array (Bower, 1984). As the clock speed increases the power dissipated by the circuitry also increases. The power dissipated by a CCD sensor also depends on the electrode structure. A two-phase device has a much larger power budget than a multiphase sensor at lower frequencies (Table 57). This extra power is needed to supply the energy required to lift the charge packets over the potential barriers in the substrate of the two-phase device. As the frequency increases, the power dissipated in the multiphase device approaches that of the two-phase device. The control of the power dissipated in the sensor (thermally generated carriers - dark current) then places a limit on how much of the circuitry is on the sensor.

Table 57. On-chip power dissipated per pixel in two-phase and multiphase CCDs as a function of clock frequency (Barbe, 1975).

Frequency (MHz)	Two Phase (watts)	Multiphase (watts)
1	1.0×10^{-6}	1.0×10^{-8}
10	1.8×10^{-5}	1.2×10^{-6}
100	2.0×10^{-4}	8.0×10^{-5}

In some CCD sensors the amplifier has been separated from the array by extending the serial register by 50 to 100 gates. One of the original reasons for this extended serial register was to eliminate device luminescence which was believed to be due to thermal input from the amplifier to the adjacent CCD pixels. Although avalanche effects in the output amplifier and partial shorts between the electrodes in the parallel and serial registers have been shown to be the source of device luminescence (Janesick, JPL, private communication), extension of the serial register could be important for the dissipation of power for a thin back-illuminated CCD. The power dissipated by the amplifier in a thick front-illuminated CCD can be neglected because of the large heat sink. Therefore, in a thinned CCD sensor, the placement of the amplifier over a thick portion of the silicon substrate should be considered since it provides a heat sink for high-frequency operations.

6.5.6 Section summary

The largest source of noise in the vector magnetograph solid-state sensor will be the photon noise which is related to the well depth. If cooled to -20°C , dark current in most solid-state sensors can be neglected. Also, assuming that double correlated sampling can be used, the amplifier reset noise can be neglected leaving only the noise associated with the output capacitance. Because of their low output capacitances, CCD sensors have a significant advantage over CIDs when comparing amplifier noise. CIDs, on the other hand, have a significant advantage over CCDs when considering transfer noise. Of all the noise sources discussed, the transfer noise has the largest uncertainty since it depends on the charge transfer inefficiency (a function of the clocking frequency, electrode structure, and driving voltages) which has not been systematically studied in large CCD sensors.

6.6 Characteristics of the SAMEX CCD Camera

The previous sections (sections 6.4 and 6.5) reviewed the terminology and the compromises that must be made in the selection of a solid-state detector. Using the results from those sections, this section will include the radiometric results of the telescope optics and the SAMEX scientific objectives (section 6.2) to develop a camera system for the SAMEX magnetograph.

6.6.1 CID Versus CCD

Although the sales literature mentions that a 1024×1024 CID is possible, several problems would exist in the standard design. Most of the CID research has been in the development of arrays using a standard TV format. This has led to a higher pixel density for a given area. Since the largest noise source is the photon noise, smaller pixels, which have smaller well capacities, will produce lower S/N ratios for single exposures. Increasing the pixel size will increase the output capacitance

which is one of the fundamental problems with the CID sensor. The capacitance load could be reduced by segmenting the array into smaller areas which would produce short electrode address lines but the complexity of the structure would go up significantly and could seriously impact the cost of the detector system. Therefore, consideration of a CID camera system for the SAMEX magnetograph will be limited to a multi-sensor detector.

6.6.2 Multi-port versus multi-sensor

There is presently only one sensor available that has the pixel density to meet the scientific requirements. Array sizes are expected to increase as companies try to develop solid-state camera systems that will compete with film cameras. Most manufacturers are reducing the pixel size as they extend the array size. For the SAMEX magnetograph the goal is to collect 10^8 electrons in 1 minute (which gives a S/N of 10^4 neglecting all noise sources except photon noise). This can be accomplished by having a "super" pixel $\sim 324 \times 324$ microns² (an unrealistic size for an array with a large number of pixels) or by adding successive images. The smaller the pixel the more images that must be added together; but more importantly, the smaller the pixels the shorter the exposure time. For high time resolution studies where the camera is reading out the data at the same time it is integrating up the next image, a shorter exposure time requires a faster readout time. Due to the uncertainties in the charge transfer characteristics of large arrays at high readout rates, a number of output ports on a single sensor may be required to reduce those rates to frequencies that are compatible with high charge transfer efficiencies. If a single sensor requires multiple ports to obtain a high S/N ratio without compromising the high time resolution, one must consider as an alternative solution the use of multiple sensors.

A multiple sensor camera system has several advantages. First, the technology for smaller arrays has already been developed and these sensors are available commercially. Also, the failure of one sensor would not represent the failure of the instrument. A disadvantage is a more complicated optical system (assuming that the sensors cannot be butted together) requiring additional relay optics for a multiple sensor camera. Careful calibration would be required to minimize any image distortion caused by differences in the relay optics.

A single sensor with multiple output ports represents a much simpler optical design. The multiple port sensor is not a new idea and requires no technological breakthrough. It has been used in high-speed cameras to decrease the clock rate (Akiyama et al., 1986) and in cameras that need to isolate signal (red, blue, and green signals in color cameras). Multiple port sensors have the same redundancy in output control circuitry as the multiple sensor camera. Although the sensor would represent a critical element, the failure of an output port (output amplifier, A/D converter, etc.) would not be critical on a multi-phase sensor

since the clock timing on the serial register could be changed so that the signal is moved to a different output port.

6.6.3 Comparison of available sensors

Figures 59-62 show how the S/N varies with temperature for four solid-state sensors. Although other sensors could have been selected, these sensors are relatively large arrays that have been developed using different approaches in optical radiation detection. For example, the Tektronix 2048 x 2048 is a three-phase, back-illuminated sensor, the TI 800 x 800 is virtual phase, front-illuminated, the NEC 960 x 1280 is a three-phase, front-illuminated interline transfer sensor, and the GE 776 x 512 is a CID sensor. In this comparison we will assume that the TI and Tektronix sensors can be operated in a frame transfer mode. As indicated earlier, if the sensors are cooled below -20 °C, the dark current can be neglected. At first glance the Tektronix array, which has the highest S/N, would appear to be the best candidate for the magnetograph detector system. This is primarily due to the large number of electrons required to saturate the 27 micron pixels. Table 58 summarizes the noise associated with the different solid-state sensors assuming a temperature of -20 °C. As discussed earlier the photon noise is the largest noise source while the second largest is transfer noise for CCDs and amplifier noise for CIDs. Even with the large saturation level, the Tektronix sensor will require image enhancement to meet the 10^{-4} polarization resolution. Therefore, we will define the enhanced S/N to be

$$S/N_{\text{enh}} = S/N (386 \text{ msec}/t)^{1/2}$$

where t is the time to reach 90% of well saturation (msec).

Table 58. Summary of the noise sources and S/N of four solid-state sensors assuming a temperature of -20 °C.

Manufacturer	Noise Sources				S/N	Exp. (msec)	Enh. S/N	Readout (pixels/sec)
	N_p	N_{dc}	N_a	N_{tl}				
Tektronix ⁽¹⁾	836	71	2	371	761	386	761	1.35×10^6
TI ⁽²⁾	361	9	3	204	314	90	650	3.54×10^6
NEC ⁽³⁾	201	26	4	215	201	145	329	4.32×10^6
GE ⁽⁴⁾	429	17	296	0	429	154	681	2.58×10^6

- (1) Tektronix, 2048 x 2048: sensor split into four quadrants (1024 x 1024) to improve charge transfer efficiency by reducing readout rates (see Figure 59). A single output port would require a readout rate of 5.433×10^6 .
- (2) Texas Instrument, 800 x 800: will require four CCDs in a multi-sensor camera to obtain the spatial resolution required by the scientific objectives (see Figure 60).
- (3) NEC, 960 x 1280: will require four CCDs in a multi-sensor camera to obtain the spatial resolution required by the scientific objectives (see Figure 61).
- (4) GE, 776 x 512: will require four CCDs in a multi-sensor camera to obtain the spatial resolution required by the scientific objectives (see Figure 62).

When comparing the enhanced S/N, the interline-transfer camera (NEC) has the lowest S/N which is due primarily to the low quantum efficiency. The CID (GE) sensor and the virtual phase sensor (TI) have enhanced S/N ratios that are comparable to the Tektronix sensor. For the TI frame-transfer sensor the enhanced S/N would represent the upper limit since it neglects the time required to shift the image from the integration area to the readout area.

Table 59 summarizes the characteristics of the individual sensors as they relate to the vector magnetograph camera design (note that availability is not included at this time although it is an important consideration). Due to the simple optical interface the next section will concentrate on the implementation of a Tektronix sensor into the vector magnetograph detector design.

Table 59. Characteristics that the vector magnetograph camera system should have.

Description	Tektronix	TI	NEC	GE
Simple Optical Interface	x			
High S/N	x	x		x
Low Readout Rate	(1)	(1)	(1)	x
No Image Smear			x	x

(1) The readout rate depends on the saturation time and the charge transfer efficiency and can be adjusted by the number of output ports on the sensor.

6.6.4 Proposed detector system

Figure 64 shows a schematic diagram of the proposed detector system. The detector sensor is a Tektronix 2048 x 2048 with a modified structure so that the readout time would match the exposure time. The array is back-illuminated with antireflection coatings to improve the quantum efficiency. The glass support structure will be mounted to the electrode ("front") side of the array so that monochromatic fringing can be eliminated. The array has been split into four 512 x 2048 areas. The two 512 segments in the center of the array represent the imaging area, and the readout areas, which are covered by a metal mask, are the top and bottom 512 segments. By segmenting the image array into two 512 x 2048 areas, the time to shift the array from the image area to the readout area has been reduced by half. The rapid shifting of the image under the masked array is necessary if the camera is to be operated without a mechanical shutter. Even with the segmentation of the array, improvements in the conductivity of the parallel gate structure will be needed to reduce the shift time from approximately 10 msec (20 msec with no segmentation, 1024 x ~20 microseconds; see section 6.5.4) to 1 msec.

The serial register, which collects the signal from the parallel transfer of the readout area, is split into two 1024 transfer registers so that the signal can be transferred to an output amplifier at each corner of the array. The readout area for each output amplifier is 512 x 1024. If each pixel in the readout area is read out in 0.7 microseconds (total time to read out the serial register is 720 microseconds), then the readout time is 367 msec which is less than the exposure time (Table 60). This decreases the processing time by allowing an image to be collected while the previous image is being read out. To

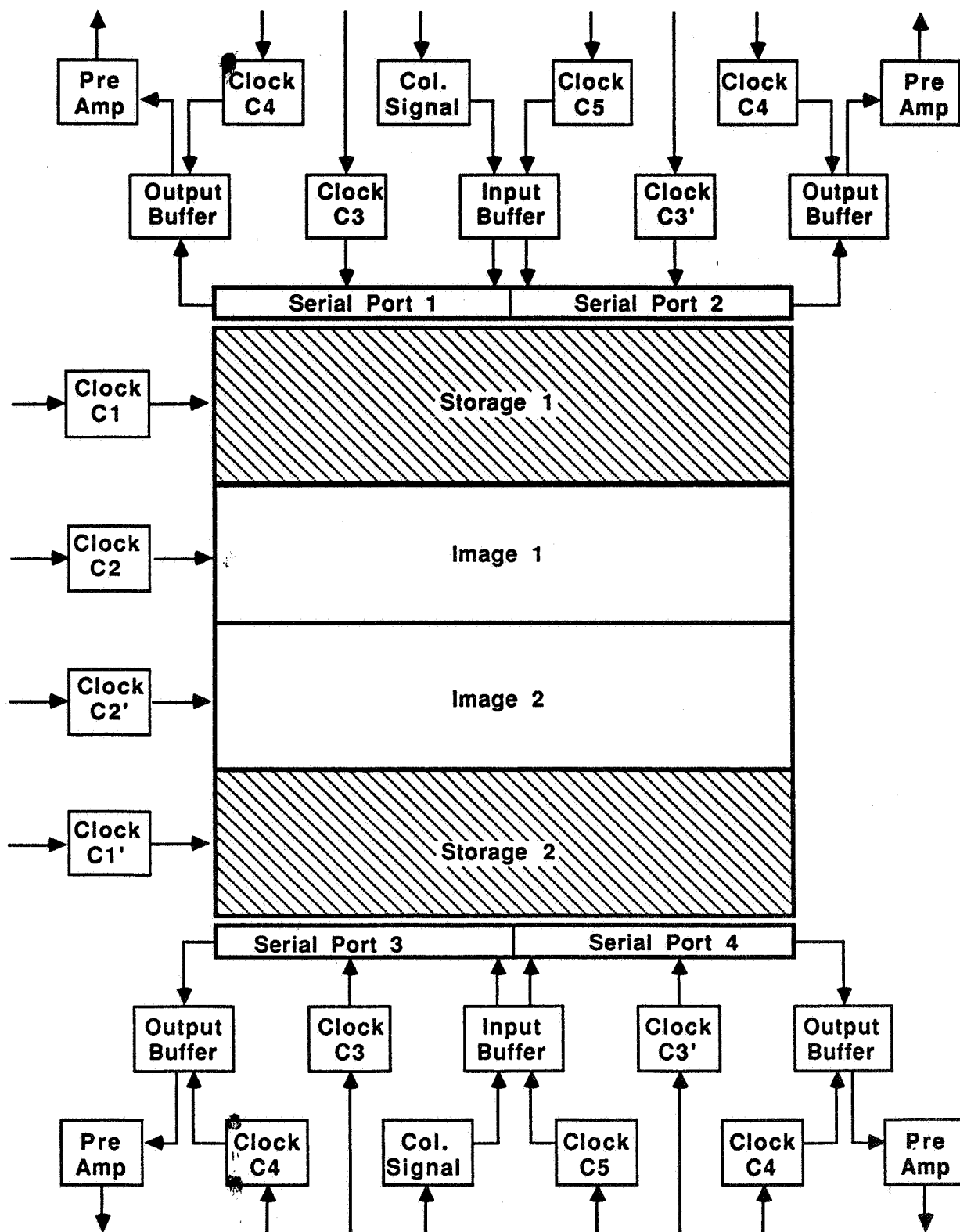


Figure 64. Proposed SAMEX detector system. This figure is a schematic diagram of a frame-transfer CCD detector system with quadrant readout to decrease the overall readout time and thus achieve the desired temporal resolution of the SAMEX magnetograph.

guarantee that the exposure time is always larger than the readout time, a polarizer following the filter will control the number of photons incident on the sensor. The signal from the four output amplifiers is then converted into a digital signal using 12-bit A/D converters. Estimated timings for the SAMEX optical system coupled with this detector array are given in Table 61. A summary of the proposed modifications to the Tektronix 2048 x 2048 array is listed in Table 62.

Table 60. Calculation of exposure time for the Tektronix 2048 x 2048 sensor.

Photons/sec on 27 x 27 micron pixel	= 4×10^6
Photons/sec on collectable area ($21 \times 27 \mu^2$)	= 3.2×10^6
Electrons/sec per pixel collected ($\eta = 50\%$)	= 1.6×10^6
Pixel well capacity (electrons)	= 7×10^5
Exposure time assuming 83% well capacity	= 0.37 sec

Table 61. Overall characteristics of the SAMEX detector system.

Exposure time	370 msec
Readout time per pixel (for each port)	0.7 μ sec
Shift time	1 msec
Acquisition time for a vector magnetogram with signal/noise = 10^4	290 sec

Table 62. Summary of the modifications required on the Tektronix 2048 x 2048 array.

1. Modification: Segment array structure into four regions

Purpose: To split the array into four 512 x 2048 parallel transfer regions. The top and bottom areas would be covered and would represent the readout areas. The two center areas would represent the imaging area.

Physical Impact: This would change the present Tektronix sensor design from a staring sensor to a (split) frame-transfer design. Three clock signals are required in the staring design; twelve would be required in a split frame-transfer design.

2. Modification: Metal or high conductivity polysilicon parallel shift gate

Purpose: To decrease the parallel transfer time from the image area to the readout area.

Physical Impact: Metal gates represent a change in the manufacturing process and would result in an increased development cost.

3. Modification: Split serial ports

Purpose: To decrease the readout time to match the exposure time so image integration and image readout can occur in parallel.

Physical Impact: If directional control of the signal packets is required for redundancy, three additional clock signals (total of six) would be needed.

4. Modification: Four input ports

Purpose: To calibrate the output electronics for each of the output ports.

Physical Impact: Would have a small effect on the physical layout of the sensor.

A mechanical shutter will be used during on-orbit calibrations to determine the effects of smearing on the image due to the split-frame transfers. The correction for image smearing becomes more critical with long parallel shift times ($>1\%$ of exposure time) which is dependent on the gate conductivity. Multiple shutters would be used and, in the case of a shutter failure, the shutter would be left in the open position or removed from the beam with a translation table. One of the shutters should be used only in parallel shift calibrations. An exit polarizer, placed between the filter and camera lens, would be used to adjust the number of photons reaching the CCD.

The camera system will include a capability of binning $n \times n$ pixels in the CCD array to improve the S/N. A reduction to an array of 512×1024 gives an effective pixel size of 0.5 by 0.5 arc sec but increases the S/N by 2 and decreases the data rate by a factor of 4.

6.6.5 Use of smaller arrays

Only one manufacturer is presently developing a sensor whose size is greater than 2000×2000 pixels. This could present a problem on the availability of a sensor for SAMEX (assuming no special funding for sensor development and if SAMEX is to fly by the next solar maximum). Certainly, the use of a multiple sensor detector system provides a viable alternative since modifications to the Tektronix 2048 \times 2048 are required to obtain the readout time, charge transfer efficiency, and the signal to noise to meet the scientific goals for the instrument. Another approach is the use of the image motion compensator (IMC) to displace the image on the sensor by a quarter of an arc second in both the x and y directions. This would simplify the optics required in a multiple sensor camera system while increasing the number of manufacturers that could develop the detector system. The disadvantage to this approach is the loss in time resolution (which could impact the scientific objectives) required to build up four images that would cover the same area as the 2048×2048 array (see Table 63). If time resolution is important, the IMC could be used to shift the image between the number 1 and 3 positions (defined in Table 63) with only a small loss in spatial resolution (~ 0.35 arc sec).

Although the IMC would represent the easiest implementation of an x,y displacement scheme, the performance would be difficult to monitor and would require calibration using solar features (sunspots). Using features that are under study for spatial calibrations is not a satisfactory approach; therefore, an extension of the x,y displacement scheme is the use of piezoelectric cells mounted on the sensor. In this approach an aperture stop (at the prime focus of the telescope) which is smaller than the array or a spatial pattern mounted in the aperture stop could be used to calibrate the x,y displacement. The image created by the piezoelectric cells could then be used

to calibrate the IMC which would provide a backup mechanism for the x,y displacement technique.

Table 63. Description of the position definition and the (x,y) displacement for each position. Movement of a 1024 x 1024 sensor to the four positions would build up an image equivalent to a 2048 x 2048 sensor.

Position #	(x,y) Displacement (arc seconds)
1	(0.00,0.00)
2	(0.25,0.00)
3	(0.25,0.25)
4	(0.00,0.25)

6.6.6 Calibration modes

Careful calibration of the detector while it is in orbit is essential if the high quality of the magnetic data is to be maintained. The detector calibrations should include:

1. Maps showing variations in the quantum efficiency over the sensor.
2. The determination of the coefficients required to correct for image smearing in a frame-transfer camera.
3. Using the input gate on a multiple port camera, the linearity and gain should be monitored for each output port. The intensity control polarizer (following the filter) can also be used if non-uniformities in the other optical components (primarily the filter) can be corrected.
4. For CCD cameras, the charge transfer inefficiencies should be monitored. The input gate on the serial register should be used to monitor the serial transfer efficiencies and a rotating slit, inserted at the prime focus, should be used to monitor the parallel registers. Using these inputs, small adjustments of the clock voltages could be made to optimize the charge transfer efficiencies.
5. For a multi-sensor camera a two-dimensional pattern should be inserted at the prime focus to calibrate out any image displacement and rotation due to optical or mechanical misalignment.

6. For a single 1024 x 1024 sensor using an x,y displacement mechanism to build up a larger imaging area, a spatial pattern in the corners of the aperture stop mounted at the prime focus of the telescope should be used to calibrate the interlacing of the images.

The first two calibrations, the quantum efficiency map and the image smear coefficients, should not vary significantly from ground-based testing and should be used to monitor the camera system. Calibrations 3 and 4, the linearity and gain of the output ports and the charge transfer inefficiencies of the parallel and serial ports, could vary as the components age and should be a part of a regular calibration sequence. The last two items are special calibrations that are dependent on the camera system selected and would not be necessary in a 2048 x 2048 array.

6.6.7 Section summary

In the previous section we have identified the modifications required to utilize a Tektronix 2048 x 2048 sensor in the vector magnetograph. Although this sensor is not currently available, the technology to develop it exists. The alternative is to design a multi-sensor detector. The industry will probably continue to develop larger arrays because of their commercial value but the development to date has been to decrease the pixel size. Large pixel, large array sensors have special charge transfer problems, but the development of such a detector is expected to continue due to the interest shown by the scientific community in such a sensor.

7. Overall System Performance

7.1 Introduction

The requirements that the SAMEX scientific objectives imposed on the design of the vector magnetograph were stringent and presented a challenging task in all aspects of the design effort. We feel we have met the challenges to produce a state-of-the-art design that incorporates significant breakthroughs in several areas of technology. The following sections summarize the expected performance capabilities of the instrument design that is the result of the different breakthroughs we have made.

7.2 Spatial Resolution

The basic requirement for a spatial resolution of 0.5 arc sec has been met. Using modern computer codes for optical design, we have achieved a design for the SAMEX optics that gives a nearly diffraction-limited image over the entire field of view. An analysis of aberration, distortion, spot diagrams, and the modulation function for the system indicates that the optical design requirements are satisfied without diminishing the overall spatial resolution. The resolution will not be impaired by image motion since any motion can be removed using a tested and/or flight-proven image motion compensator.

7.3 Field of View

The field of view we have chosen is 4.3 x 8.5 arc min. This field can envelop the area of most active regions which are the primary targets of the SAMEX instruments. While full-disk magnetogram images are desirable, they are not feasible with the SAMEX instrument if we are to achieve a spatial resolution of 0.5 arc sec. However, the capability for acquiring full-disk magnetograms can be recovered through a mapping of the disk.

7.4 Magnetic Sensitivity

We have shown that it is possible to achieve a sensitivity of 30 to 50 G in measurements of the transverse component of the magnetic field with a time resolution of less than 5 minutes by using CCD detectors and a process of image enhancement. However, achieving this sensitivity is possible only if the overall sensitivity of the foreoptics and polarimeter is at least 10^{-4} . The signal-to-noise ratio (S/N) of CCD chips is on the order of 400 to 600, but the noise is for the most part stochastic, governed by Poisson statistics, and can be reduced through the process of adding images together - image enhancement. On the other hand, instrumental polarization introduced by the optics and the instrumental noise of the polarimeter cannot be reduced through image enhancement. This means that the accuracy of the polarimeter must be equivalent to the desired sensitivity (10^{-4}) and the instrumental polarization must be at least an order of magnitude less. Both these conditions posed difficult

technological challenges. An extensive analysis of different techniques for polarimetry pointed the way to a polarimeter that could measure linear polarization to an accuracy of 10^{-4} : a rotating hybrid polarizer composed of a Glan-Thompson prism and a HN38 sheet polarizer. However, to achieve this accuracy in the polarimetric measurements, we could not use conventional optics in the design of the foreoptics since they would introduce instrumental polarization on the order of 10^{-3} that would outweigh the accurate measurements of the polarimeter. We required the instrumental polarization introduced by the foreoptics be less than 10^{-5} and this condition compelled us to apply the newly developed technique of polarization aberrations to the foreoptics design. Using this technique, we have derived a system of optical coatings for the two mirrors of the Cassegrain and for all the relay lenses so that the entire optical system of the foreoptics produces less than 10^{-6} instrumental polarization. Because of these two advances in technology, we expect to achieve the required magnetic sensitivity with the SAMEX instrument.

7.5 Temporal Resolution

Important advances had to be made in other areas to define the temporal resolution of the magnetograph. We had to specify a large-array CCD camera system to meet the requirements on the field of view, but we also specified that the camera have multiple readout ports to keep the readout time below the exposure time. The camera also had to use a frame-transfer sensor so that half the sensor is used for imaging and half for storage of an image while the image is being read out. Such a camera would incorporate several technologies that have not been combined before now. Based on the assumption that this type camera can be made, we defined the temporal resolution of the instrument based on the estimated exposure time. This estimation is uncertain to some degree because it depends on how closely we can estimate the transmission characteristics of the optics. If we take the derived value of 0.37 second for this exposure time, we can show that measurements of the vector magnetic field can be made in less than 5 minutes with an accuracy of 30 to 50 G in the transverse component and 1 G in the line-of-sight component.

In summary, then, we expect to achieve the required magnetic sensitivity in less than 5 minutes with a spatial resolution of 0.5 arc sec over an optimum field of view. Indeed, we have succeeded in designing a state-of-the-art instrument for the SAMEX vector magnetograph.

III. SPACECRAFT REQUIREMENTS

1. Onboard Computer System

1.1 Introduction

The SAMEX vector magnetograph will need a sophisticated onboard computer system to control the operation of the instrument, to acquire data, to perform data reduction, analysis and storage, and to communicate with the host spacecraft's computer. It must also transmit and retrieve data from the NASA Tracking Data and Relay Satellite Systems (TDRSS) via the spacecraft host computer.

The current (1986) space-based computer technology is in the PDP 11/34 class with executions of some 120,000 Fortran instructions per second and 250,000 Wheatstone instructions per second with power requirements of 15 W and a weight of 15 kg (Hogan, 1986). The memory capacity is in units of 128 kbytes of Dynamic Random-Access-Memory (DRAM) and 128 kbytes of Erasable Programmable Read-Only-Memory (EPROM) of CMOS chips (Complementary Metal-On-Silicon field effect transistor). An 8206 error correction code unit can be used to detect and record single-event data upsets caused by cosmic ray interaction with the memory (Hogan, 1986). Computer technology should advance sufficiently over the next few years that we can anticipate a MicroVAX II class machine will be developed to operate under space flight conditions in the 1990's, the era of Space Station technology. As an example of technological advances, current dynamic random-access memory (DRAM) chips are now being constructed in the laboratory with 6 million binary bits (16 Mb), and megabit chips are now being produced (Robinson, 1987).

1.2 Computer System

The computer system will consist of a control processor, data processor, Fast-Fourier Transform (FFT) processor, image-motion compensator (IMC) processor, arithmetic unit, adder-subtractor units, and interfaces with the SAMEX host computer and the H-alpha imager. It will provide data storage management and will control the thermal environment, polarimeter, filter, filter temperature, shutter, and detector.

An overview of the system is shown in Figure 65. The antenna-transponder receives commands and sends the telemetry data from the experiments through the host spacecraft's computer system. The host spacecraft will provide the location of the spacecraft in orbit so that the magnetograph's main computer can compute the heliocentric Doppler shift and correct for the shift in the wavelength position of the spectral filter. The host computer will supply the universal time, time of orbital noon,

ELECTRONIC FUNCTIONAL DESIGN

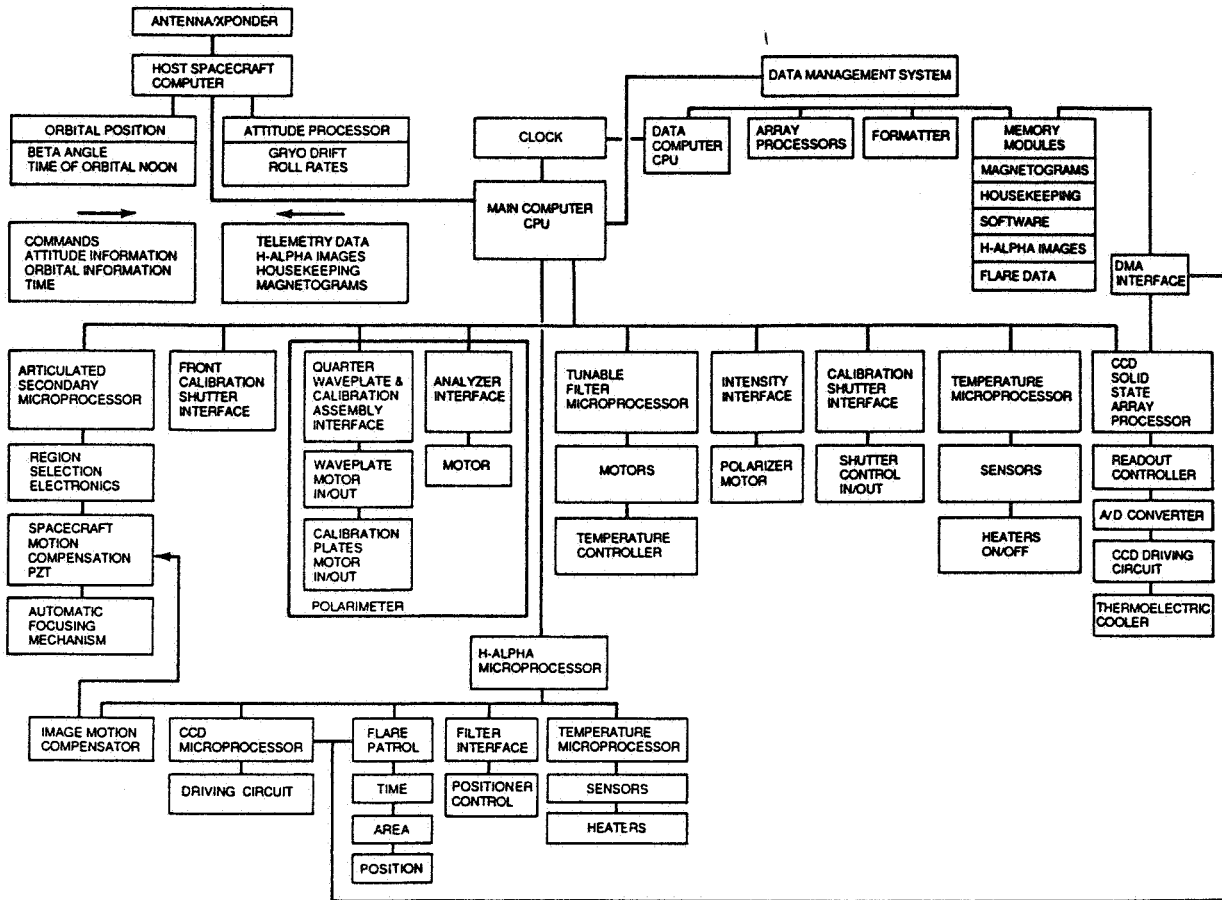


Figure 65. Functional design of the computer system for the SAMEX vector magnetograph and H-alpha instruments. Each block represents an identified function that the main computer will have to control and operate. Also shown are blocks representing the various microprocessors that will interface with the main computer. The functions of the computer for the host spacecraft that are relevant to the magnetograph/H-alpha system are also shown.

the orbital period, the orbital beta angle or inclination, and the time of ascending node. The heliocentric Doppler velocity for filter corrections will then be made by the magnetograph's computer. The times of sunrise and sunset will also have to be calculated to determine if the commanding sequences can be started, finished, or continued. The host spacecraft's pointing errors, gyro drift rates, and roll rates are needed to correct for offset pointing and for servo control inputs.

The main instrument computer (MIC) for the magnetograph will be the command center to control initiation of various processes. The MIC will oversee data processing tasks and provide control over the articulated secondary microprocessor, shutters, polarimetry sequences of the quarter-wave plate and polarizer, the filter microprocessor, the intensity controller, instrument temperature regulation, the detector processor, the H-alpha microprocessor, and all interfaces with the host spacecraft.

In Figure 66 we show a preliminary block diagram of the electronics in which the interrelationships between components are displayed. Detailed structure of the electronics will depend greatly on the computer system available.

For the magnetograph, a 32-bit MicroVAX II class machine containing a high-performance floating-point (MFLOPS) array processor on a Direct Memory Access (DMA) input/output bus interface would be desirable for onboard image processing and memory requirements. A Sky Warrior 15 MFLOPS onboard array processor interfaced to a MicroVAX II's Digital Electronic Corporation Virtual-Memory-Extension bus (DEC VMSbus) is an example of a system configuration that might be considered. However, a smaller class computer could be used to handle the data. For example, the NASA NSSC-I computer, with technology circa 1972, is the central computer for the scientific instruments for the Hubble Space Telescope, and the demands on this computer are similar to those of the SAMEX vector magnetograph in the area of data rates and commands. The computer is made by IBM and is based on a 16-bit 8086 cpu processor. Another example of a system using a smaller computer is the Solar Optical Universal Polarimeter (SOUP) instrument flown in July 1985 on Spacelab 2. This instrument used an LSI-11 microcomputer (technology circa 1980) with a video processor to provide control and operation of the instrument and to handle the data from a small detector array of 250 x 250 pixels (Lingren and Tarbell, 1981). A memory of 10 Mb combined with the video processor based on Advance Micro Devices' 2900 family of bipolar bit-slice devices provided the array processor for the experiment. But as technology advances, by the late 1980's industry should be able to provide a 32-bit MicroVAX II class computer for the main computer of the SAMEX magnetograph.

In Table 64 we list the general requirements of the computer system(s) for the SAMEX vector magnetograph. These are based on the following operational parameters:

BLOCK DIAGRAM OF ELECTRONICS

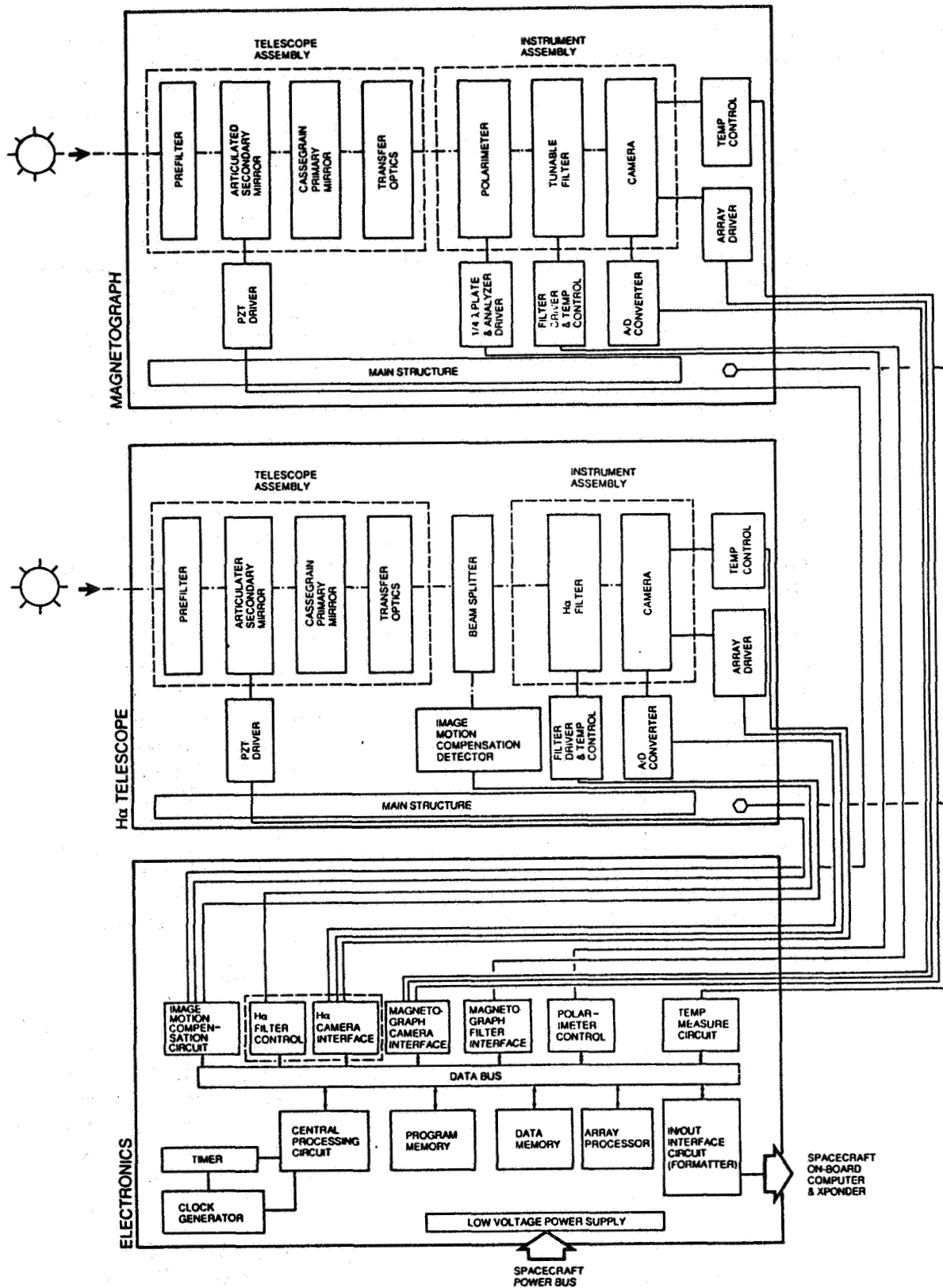


Figure 66. Block diagram of electronic interfaces identified for the magnetograph and H-alpha instruments.

- single polarization image from CCD A/D: 12 bits x 1024 x 2048;
- time to acquire a single polarization image: 0.372 sec;
- enhanced circular polarization image (50 additions): 18 bits x 1024 x 2048;
- enhanced linear polarization image (225 additions): 20 bits x 1024 x 2048;
- storage for 1 vector magnetogram (I_V , I_Q , Q , U , V): 14 bits x 5 x 1024 x 2048.

Table 64. Computer system requirements.

Adder Memory

Word size: 20 bits
Memory: 42 Mbits (20 bits x 1024 x 2048)

Mass Storage Memory

Word size: 14 bits
Memory: 1.5 Gbits (10 vector magnetograms = 10 x 5 x 14 x 1024 x 2048)

Array Processors

Floating-point array processor for polarimetric calculations of the Stokes parameters and FFT potential calculations

Internal Data Rates

1. From CCD A/D to adder: 68 Mbps (12 bits x 1024 x 2048/0.372 sec)
2. From Adder to disk: 2 Mbps for circular polarization images (18 bits x 1024 x 2048/18.6 sec)
0.5 Mbps for linear polarization images (20 bits x 1024 x 2048/83.7 sec)

Downlink Telemetry Rates

TDRSS allowable rates: 30 Kbps to 3 Mbps
1 vector magnetogram: 0.15 Gbits in 5 minutes = 0.5 Mbps
Time to downlink 1 vector set @ 30 Kbps = 82 min
Time to downlink 1 vector set @ 3 Mbps = 50 sec

As part of a NASA Small Business Innovative Research (SBIR) program, BGB, Inc. is currently designing a computer system for MSFC that would be used for a space-based vector magnetograph. When completed, this design will give us a better concept of the type of computer system that can be developed with present-day hardware.

1.3 Microprocessors

Microprocessors will be used to control the many subsystems of the SAMEX magnetograph and H-alpha telescope: solid-state camera, optical filter, thermal control systems, polarimeter, telescope pointing and guiding, and array processing. These microprocessors will reduce the load on the main computer system(s).

The microprocessor for the articulated secondary mirror will have to interface with the MIC and the detector system of the H-alpha image motion compensator. The off-set pointing requirements and the analog-digital signals from the fine guidance subsystem will create servo instructions for the piezoelectric drives of the secondary mirrors of the H-alpha and magnetograph telescopes and their readout transducers (position encoders).

The microprocessor for the polarimeter receives commands for position sequences for the quarter-wave plate and polarizer, and also generates the corresponding commands to the motors that position these optics and verifies the completion of the positioning movements.

The spectral filter's microprocessor controls the tuning of the filter to the appropriate line and line offset, continuously corrects for the heliocentric doppler shift as the orbital velocity changes, and regulates the temperature of the filter components.

The detector microprocessor controls the readout controller of the CCD solid-state array, the CCD driving circuit, binning of pixels, and thermal control of the thermoelectric cooler.

The microprocessor for the H-alpha telescope controls the detector package and the filter and is used to analyze the H-alpha images for flare activity by monitoring brightness enhancements in the images.

1.4 Mass Storage

Because of limitations imposed by available downlink telemetry using the NASA TDRSS, there will have to be mass storage capabilities associated with the SAMEX magnetograph's main computer to hold and process the raw data from the magnetograph. As indicated previously, storage of ten processed vector magnetograms will require about 1.5 Gbits of memory.

CMOS/SOS (complementary-symmetric metal oxide on silicon/silicon on sapphire field effect transistor) integrated circuits with memory of ~1 Gbit can be used since they satisfy the memory requirements, require low power, and are immune to cosmic ray upsets because they have no PNP junctions. An error correction code can be used to correct single event errors due to single cosmic ray events (Hogan, 1986).

The accumulating memory requirement is similar in size to the STARLAB project studied by Mount Stromlo and Siding Spring Observatories and Matra Space in 1981 (Mount Stromlo and Siding Spring, 1981). The memory technology (Borsky, 1981) available then was 256 DRAMS on chip carriers in a 22 x 22 x 35 cm³ package weighing 9 kg and using 228 watts of power. In 1986 the technology (e.g., Dataram Corporation) could provide 16 Mbyte of DRAM on a board for a ground-based VAX 8600/8650 system. The needed ~1 Gbit of memory should be available when the SAMEX magnetograph is developed because of rapid development in this technology. Industry is currently developing a 500 Mb memory module and a 1 Gb (10⁹) board is also on the horizon with a weight of 4 kg and a power requirement of 15 W.

The Japanese Rotary Digital Audio Recording Technology (R-DAT) can provide a 1.38 Gbyte storage cassette in a 7.3 x 5.4 x 1.05 cm³ device. The development in memory technology will apparently meet the needs of the SAMEX magnetograph system.

1.5 Software

The onboard data analysis programs would include: (1) instrument checkout routines, (2) multi-experiment interactions for start of operations of other experiments and for reading data obtained from other experiments to start a magnetograph sequence (e.g., flare mode), (3) potential field calculations, and (4) H-alpha data analysis to determine locations of flaring activity and timings of flare onset, maximum and decay, with reference to starting the SAMEX magnetograph in a flare sequence in which magnetograms are acquired rapidly and stored for later analysis and/or transmission. All of these programs can be tested at the MSFC vector magnetograph facility on microprocessors and central controllers. A large set of software has been developed for the MSFC vector magnetograph and it will be available for adaptation to SAMEX requirements. Table 65 gives a listing of some of the operational software available at MSFC.

Table 65. Operational software for the MSFC vector magnetograph

Operational Mode	Programs
System diagnostics	Microprocessor testing Bad pixel/locate Bad pixel/correct System interrupt
System setup	CCD exposure time Centering of filter wavelength
Data acquisition	Automatic vector sequence Selected sequence (B_L or B_T) Automatic filter scan
Calibration	Offset guider readout Polarimeter calibration Detector calibration/zero illumination Detector calibration/flat field Instrumental polarization
Analysis	Polarization cross talk Signal-to-noise Image histogram intensity plot
Files management	Copy files to tape Delete files

The selection of onboard software will both drive and be driven by the design of the electronic computer.

1.6 Telemetry

The onboard computer system must interact with the spacecraft telemetry system to transmit the SAMEX data to the ground station. The details of the requirements placed on the data rates by TDRSS are given in section 5.

2. H-Alpha System and Image Motion Compensation

2.1 Introduction

The H-alpha telescope will provide: (1) imaging of the tracers of the chromospheric magnetic field (fibrils, filaments, and arch filament systems); (2) chromospheric observations of flares and filament activity; (3) cospatial identification of optical emission from magnetic features seen in magnetograms (e.g., network filigree, Ellerman bombs, locations of umbrae and penumbrae); (4) chromospheric Doppler velocities and sunspot motions; and (5) a full-Sun image that is diverted to the

detector in the image motion compensator system for corrections of image motion. To accomplish these objectives, a telescope with the same spatial resolution as the magnetograph's telescope is required, along with a narrow-band, tunable H-alpha filter.

2.2 Optical System and Detector

The optical design for the H-alpha telescope is shown in Figure 67. The basic elements are a Cassegrain telescope, tunable H-alpha filter, image detector, and the sensor system for the image motion compensation. The telescope is an aplanatic Cassegrain design, Ritchey-Chretien, which provides minimum spherical aberration and coma, and provides the shortest telescope for the longest effective focal length. The telescope has the same characteristics as the magnetograph's telescope to obtain the same spatial resolution. We require separate optics for the H-alpha imaging to avoid introducing any polarization into the magnetograph optics with a beam splitter and to achieve a shorter overall system. In addition, by having a separate H-alpha telescope, the prefilter for the magnetograph can be narrower, thereby reducing the heat flux introduced into that system. To obtain H-alpha images with the magnetograph, the bandpass of the prefilter would have to be widened an additional 1000 Å. The filter's spectral bandpass of 0.5 Å is selected to minimize the effect of the orbital velocity of ± 7 km/sec (± 0.125 Å) with respect to the Sun and to be compatible with the full-width-at-half-maximum of 1 Å for the 6250 Å H-alpha line (Gibson, 1973). The relatively broad 0.5 Å width of the H-alpha filter allows various tunable filters to be considered.

The CCD detector for the H-alpha telescope system is identical to that for the magnetograph. This will allow for simplification of design and possible interchange of the detectors for testing and optimization of the magnetograph measurements. It also provides redundancy for the magnetograph system.

2.3 Image Motion Compensation

Since the total spatial resolution achieved by the system depends on both the system's optical resolution and the stability of the image at the detector, the image motion compensation system (IMC) must provide pointing stability to within 0.1 arc sec for the magnetograph as shown by Tarbell et al. (1981). There are various ways to design an IMC system. The methods described below use the limb of the Sun as the reference for the line-of-sight pointing corrections. A beam splitter before the focal plane of the H-alpha Cassegrain telescope would provide an image for the IMC (see Figure 68a). The beam splitter would be a notched interference filter that feeds an image of the photosphere seen in the off-band H-alpha continuum to the IMC, while passing an image (H-alpha ± 10 Å) to the aperture stop for the H-alpha imaging system. This method of deflecting a photospheric image to the IMC will not produce any effect on the

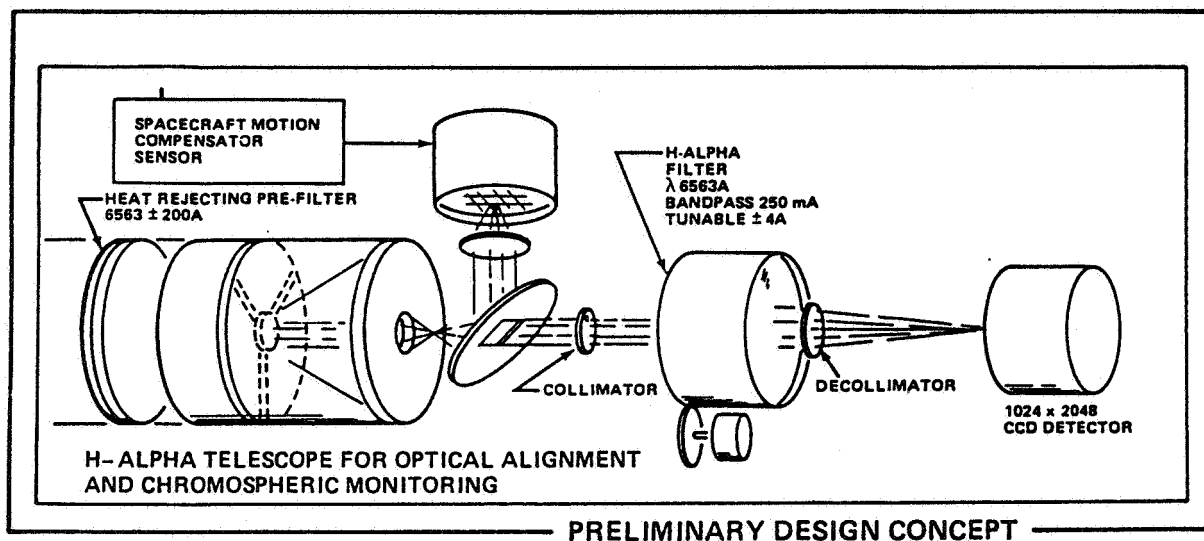


Figure 67. Schematic design for the SAMEX H-alpha telescope system. The basic components are a 30-cm Cassegrain telescope identical to the telescope for the SAMEX vector magnetograph, a narrow-band, tunable H-alpha filter, and a CCD camera system identical to the detector for the magnetograph. Also shown is the optical system that deflects the full solar image into the sensor for the image motion compensation system.

field of view of the H-alpha imaging system. Any solar limb motion is detected at the IMC's detector, and error signals are then input to a feed-back loop which controls the compensating movements of both the magnetograph and H-alpha secondary mirrors. Since the exposure times needed to obtain H-alpha images are relatively short (~1 sec), it would not ordinarily be necessary to use image motion compensation for the H-alpha observations. However, since we require extremely accurate cospatial registration of the H-alpha images and magnetograms, it will be necessary to impose identical compensating motions on the secondary mirror of the H-alpha telescope. To have a closed-loop system, the secondaries should have readout transducers to insure that the motions of the secondary of the magnetograph telescope are following those of the H-alpha telescope.

The error signals for image motion compensation are provided by a dual-quadrant photodetector system positioned on a full-Sun image obtained from the H-alpha telescope. Two sets of four detectors positioned at 45° relative to each other provide two redundant systems. Since the linear detectors cannot interlock at the center (see Figure 68), the gap there is covered by a small, square solid-state array. This array is needed for those times when the off-set pointing is to an area very near a limb of the Sun; this pointing would place one of the solar limbs in the gap at the center of the linear photodetectors. Defocusing the image of the Sun would provide a greater dynamic range for positioning the solar limb on the detector by extending the size of the Sun's image; however, this method is not used in the system described herein.

Signals can be summed from the pixels of an array detector operating as individual quadrant detectors, or an integrated system can be used. If photodiode detectors are used to integrate the signals, then an electrical integration network of analog amplifiers and dividers can be used to determine the center of the solar disk. The first stage of the network would boost the photodetector current and convert it to a voltage, and the second stage would perform the summing and differencing of the various quadrant signals. If we define the output signals from the four linear photodetectors as A, B, C, and D as illustrated in Figure 68, then the X and Y positions of the solar image are given by:

$$X = k[(B+D)-(A+C)]/(A+B+C+D)$$

and

$$Y = k[(A+B)-(C+D)]/(A+B+C+D),$$

where k is an instrumental calibration coefficient that converts voltages to position coordinates.

For an H-alpha telescope with a 100-mm image of the Sun and a two-dimensional detector array with 27-micron pixels, the 1800 arc sec image of the Sun could be located to within 0.48 arc sec of the limb, i.e. (1800") x (0.027/100). The array would allow the limb to be positioned to a much higher accuracy than the 0.48 arc sec resolution of a single pixel by using the

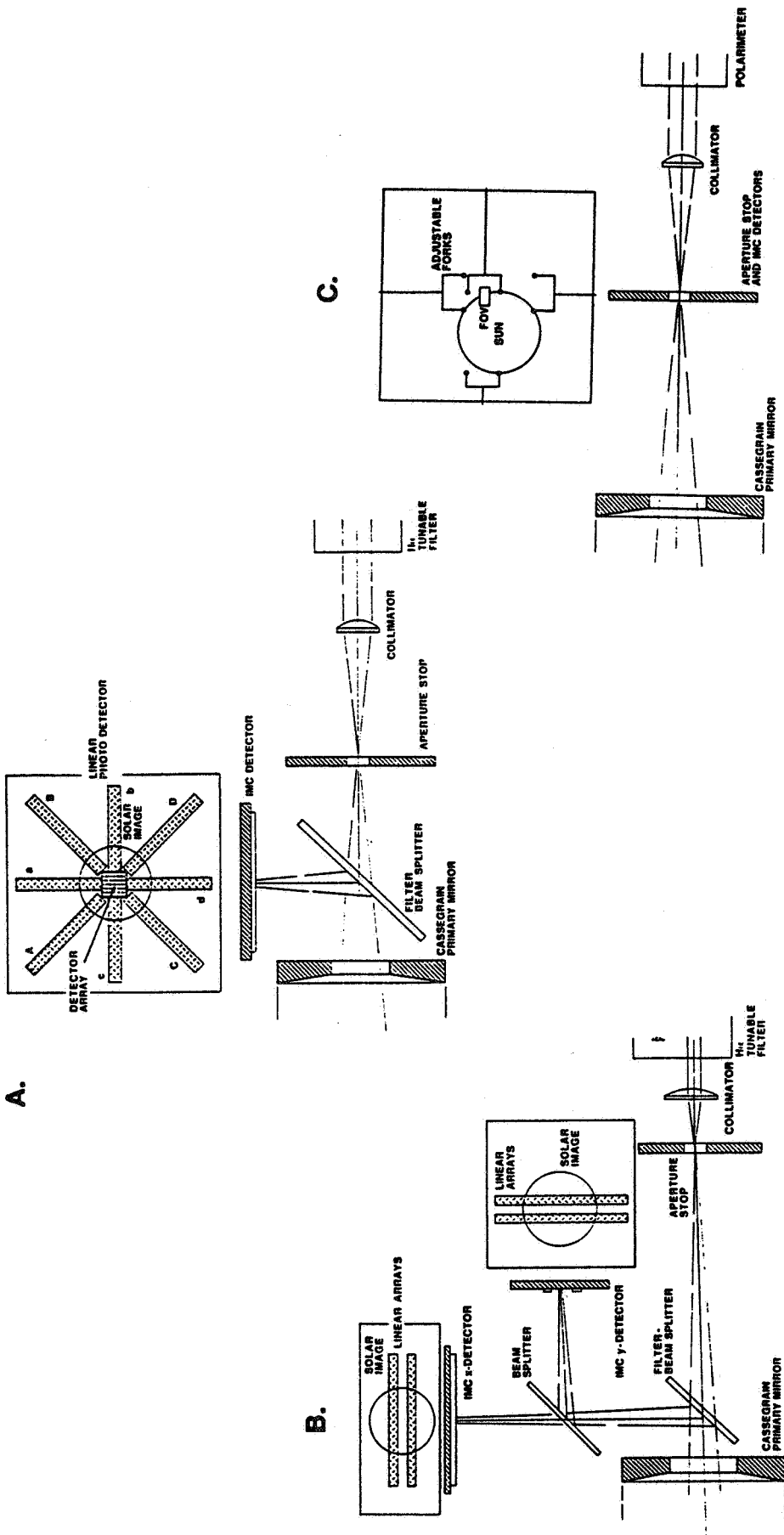


Figure 68. Schematic designs for the image motion compensation (IMC) system. Shown are two possible designs that could be coupled with the H-alpha telescope optics and one design coupled with the magnetograph optics. In all three designs, the system will provide image motion compensation for both the H-alpha and magnetograph instruments. The basic idea in all three designs is to detect relative motion of the solar limb in two axes and to correct for this motion by moving the secondary mirrors of both telescopes. (a) IMC using solar image from the H-alpha telescope. A notched interference filter feeds an image of the full Sun (as seen in off-band H-alpha) to a detector consisting of eight linear diode arrays plus a small, square array at the center where the linear detectors cannot interlock. (b) IMC using two solar images from the H-alpha telescope. Here the filter beam splitter feeds another beam splitter so that two images of the full Sun are created. For each image, two linear arrays are used to sense the solar limb in one of two axes. (c) IMC using solar image from the magnetograph telescope. Four double-forked detectors are located at the prime focus of the magnetograph telescope where the solar limb is imaged. This design was used in the Spacelab 2 SOUP instrument.

statistical averaging of a number of rows and columns of an array. However, since the field of view of the instrument is 4.3' x 8.5', there is no need to have the limb at the very center of the field of view. The use of the total array image will provide further accuracy of position. A centroid of the Sun given by an average in X and Y taken over an array of 1800 pixels (with a pixel resolution of 1 arc sec over the 1800 arc sec solar diameter) would have a statistical uncertainty of only 0.03 arc sec in both X and Y. We do not consider using a large single array in place of the combined linear photodetectors and small square array at the center because the temporal response of the IMC would be limited by the time taken to read out the array (although selected reading of an array could be done with a CID). Except for limb observations, only the linear detectors will be used to generate error signals. The general positioning of the field of view on the Sun depends on the calibration constant of the detectors and the electronics. Such systems have been developed to attain 0.1 arc sec positioning (Light, 1982). If the pointing accuracy is 0.25 arc sec, the sensitivity is on the order of 10^4 (0.25"/1800").

An alternate design (Figure 68b) uses two beam splitters to form two images, one for detecting motions in the X-axis and one for the Y-axis motions. The set of linear photodetectors shown interlock at the center so that the limb can be detected when the telescope is pointed to an area very near the limb. This design thus eliminates the need for a separate square array at the intersection of the quadrant design. The double linear arrays in each axis provide redundancy. If the separation of the two linear arrays is equal to the radius of the solar image, then an angle between the linear photodetectors and the solar limb of greater than 45° is guaranteed for one of the arrays (e.g., the SOUP system). Because of the large field of view in the SAMEX design (4.3' x 8.5'), the limb does not have to be exactly at the center of the field of view so the separation between detectors is not critical.

The image motion compensator for the Spacelab 2 SOUP instrument provided a 2-sigma pointing stability to less than ± 0.01 arc sec with a peak-to-peak excursion of less than 0.04 arc sec; this performance improved the pointing stability of the Spacelab pointing platform (the Spacelab Instrument Pointing System, IPS) by a factor of 500 (Title et al., 1986). This limb-tracking guidance system worked in servo with the articulated secondary mirror of the SOUP telescope. The SOUP design used a set of four double-forked detectors located at the prime focus and aperture stop for the rest of the optical system (see Figure 68c). To avoid interference of these detectors with the field of view passed by the aperture stop, the detectors could be moved in and out of a 1.1° field (Tarbell et al., 1981). In the SAMEX vector magnetograph design, there is no need for movement of the limb detectors because the image used for motion compensation is formed by a beam splitter in the H-alpha system. This will simplify the SAMEX design but still provide the needed 0.01 arc sec stability required for 0.5 arc sec spatial resolution. Since the SOUP design (Figure 68c) has no optical elements in the field

of view of the magnetograph, the system could be used to control the magnetograph directly by having the detectors at the prime focus plane of the magnetograph telescope and not on the H-alpha system as with the other above designs. This would, however, lengthen the optical design of the magnetograph and not allow the placement of a field stop and mechanical shutter at the prime focus of the instrument.

The H-alpha imaging system can be used as a two-axis recognition tracker to recognize and follow specific features in its field of view. The particular H-alpha feature, which might be in a sub-section of the total field of view of the H-alpha image, would be moving with respect to the disk due to solar differential rotation and photospheric motions. This special tracking mode would be useful for analysis of small, moving, and dynamic magnetic features such as individual granules. Tracking specific features could be accomplished by software comparison of images directly or using Fourier transform analysis of the images diverted to the image motion compensation system and used to develop the spatial corrections. The Fourier analysis method would use the complex two-dimensional Fourier coefficients a_{nm} of order (n,m) of an image at time $t = t_0$, and the Fourier coefficients, A_{nm} of an image at $t = t_0 + dt$. As an example of the method, consider the one-dimensional case in x ; then a_{nm} and A_{nm} are just a_n and A_n . Then, because of the phase shift properties of the Fourier coefficients, an expression for the spatial movement of features, Dx , between the images from time $t = t_0$ to $t = t_0 + dt$ is given by

$$Dx = i/n \ln(A_n/a_n),$$

where only one order of complex coefficients is required (CIT Photoheliograph Definition Study, Final Report, Ball Brothers Res. Corp., 1971, NAS8-30190, pp. 5-33). This procedure for feature tracking would require a system for target acquisition to bring the object to within the field of view of the H-alpha system. A two-dimensional analogy could employ the H-alpha images or selected sub-frames for rapid updates and tracking specific features. A separate limb tracker is required in any case since any correlation tracker observing a small part of the Sun will follow the solar rotation and a means of compensation for stability must be provided (Smith and Tarbell, 1977).

3. Experiment Observing Modes

3.1 Introduction

The three SAMEX instruments - vector magnetograph and H-alpha and X-EUV imagers - will have a number of unique sets of scientific objectives, each set determined by a specific scientific question. Addressing these various science objectives will require selection of an appropriate common target and definition of the various operational parameters for each of the instruments, e.g., choice of spectral line(s) and spectral position for the magnetograph's filter, choice of on-band or off-band filter position for the H-alpha images, and spectral range

and bandpass for the X-EUV instrument. Consequently, there is a need for an effective technique for planning the observing routines around each particular science objective to provide joint observing programs for the study of selected phenomena. Moreover, selected observing programs for different objectives will have to be combined effectively and efficiently to keep data rates consistent with the restricted telemetry rates of TDRSS. One technique that can be used is to place the science requirements into specific observing building blocks. This approach was used effectively in both the Skylab-ATM and Solar Maximum Mission operations to determine timelines for operations of the various experiments and for telemetry.

The primary objective of the SAMEX magnetograph is to obtain very accurate measurements of the vector magnetic field of features associated with solar activity with high spatial resolution. Through the combination of high spatial resolution with the temporal continuity that space observations allow, we will be able to investigate a variety of short-lived and long-lived features for the first time. The polarimetric sensitivity we will achieve with the SAMEX instrument will allow us to measure the important magnetic characteristics of solar phenomena to a degree never before possible. Specific scientific studies to be carried out with this unique instrument were identified in the sections on scientific objectives; they are:

- (1) active regions
- (2) flares
- (3) prominences and filaments
- (4) synoptic observations
- (5) fluxtube diagnostics

Further observing modes will have to be designed to perform necessary calibrations.

Through careful design of the various observing modes, we can obtain coordinated measurements of the photospheric vector magnetic field, photospheric and chromospheric velocities, and the morphology of the chromospheric and coronal magnetic fields. These data will provide the basis for many outstanding research programs in the years ahead - research that will be at the forefront of solar science. In the following sections, we outline some possible observing modes for these scientific studies. The modes presented are given only as illustrative examples of the observational schemes which must be developed for all the SAMEX instruments and included in the development of all the software for the onboard and ground-based computers. We do not suggest that the following observing modes are complete; many iterations will have to be made to reach the final set of observing building blocks.

3.2 Active-Region Mode

The observational programs to study active regions (AR) can be divided into the following modes:

1a AR rapidly developing	1e AR on limb
1b AR highly sheared	1f AR structure
1c AR near disk center	1g AR sunspot
1d AR off disk center	1h AR bright spots
1i AR velocity and flows	1j AR coronal correlation

These modes provide for the study of the evolution and structure of the region (1a, 1b, 1f, 1g), the dynamics and velocity fields (1b, 1i, 1h), the height extension of the regions and coronal correlations (1j), and the effect of viewing geometry (1c, 1d, 1e).

3.3 Flare Mode

Flare observations can be divided into the following modes:

2a FM magnetic shear	2e FM pre-flare activity
2b FM fast mode	2f FM loop structure
2c FM free energy	2g FM mass motion and flows
2d FM current system	2h FM coronal bright points

The flare modes study magnetic shear (2a), the rapid variation of the field (2b), the interaction of magnetic and fluid velocity fields (2d, 2f, 2g), and the evolution, source, and dynamics of a flare (2c, 2d, 2e, 2h).

3.4 Prominences

Observations of active and quiet prominences can be divided into the following modes:

3a PM magnetic neutral line	3d PM magnetic shear development
3b PM Doppler study	3e PM active prominence
3c PM filament structure	3f PM quiet prominence survey

The prominence modes study the structure and evolution of both weak and strong magnetic fields in and below prominences. The study of the detailed structure of regions of weak magnetic field is particularly suited to the SAMEX magnetograph since its state-of-the-art sensitivity will allow detection of perturbations caused by emergence of new magnetic flux and changing velocity patterns.

3.5 Synoptic Surveys

The synoptic survey of sunspots and patrol observations can be divided into the following modes:

4a SS full-Sun mapping	4f SS flux cancellation
4b SS coronal holes	4g SS weak-field granulation
4c SS polar regions	4h SS ephemeral regions
4d SS large-scale migration	4i SS flow patterns
4e SS flux emergence, submergence	

The synoptic survey modes listed here reflect the general areas of the Sun to be studied. The full-Sun mode will provide an important data set to serve as background for the entire research program.

3.6 Flux Tube Diagnostics Mode

The observing programs for flux tube diagnostics will involve extended observations to obtain Stokes data at small wavelength increments within single spectral lines and for several different lines. These data will provide the information needed to define the physical conditions that influence the formation of the spectral lines. The diagnostic modes are listed below:

5a DM line profiles	5d DM temperature effects
5b DM multiple line analysis	5e DM filling factors
5c DM pressure effects	5f DM Doppler motion

The diagnostic modes will provide the data required for an accurate analysis of the magnetograph data. The effects of temperature, pressure, filling factors, and Doppler motions on the Stokes profiles must be included in the radiative transfer theory that is the basis for interpreting the magnetograph data.

3.7 Calibration

Calibration modes will provide necessary data for monitoring instrument performance throughout the mission. The calibration modes are:

6a CM detector performance	6d CM IMC performance
6b CM polarimeter performance	6e CM shutter performance
6c CM filter performance	6f CM onboard software

The calibration modes will allow calibration of the system and system components at regular intervals. An onboard internal calibration of many of the components using "artificial intelligence" technology would be desirable to insure proper operation of the instrument without continuous monitoring of the system from ground-based stations.

3.8 Summary

These various modes of observation will translate into selections of instrument settings, observational cadences, and observed solar features. In Table 66 we list the various options that one would use in designing specific modes of observations. These provide examples of what might be loaded into the onboard computer memory to simplify the process of command generation.

Table 66. Command building of observational modes.

Solar Feature

Name of feature:	NOAA Boulder number
Heliographic coordinates:	theta, phi
Mapping mode:	on or off

Observation Cadence

Start time:	UT1
End time:	UT2
Cadence:	delta t
Flare mode:	initialization via H-alpha observation or x-ray data

Magnetograph Instrument Parameters

Filter settings:	wavelengths setting with line center and offset filter position
Polarimeter sequences:	quarter-wave plate in or out positions of polarizer
Shutter mode:	mechanical or shift mode
Intensity control	setting
Camera control:	number of enhancements array sub-set parameters (Xo, Yo, DX, DY) for readout of the selected areas

H-Alpha Instrument Parameters

H-alpha filter settings:	positions of filter
Detector control:	array sub-set parameters of readout
Flare patrol mode:	brightness analysis criteria recording mode

4. Spacecraft Options

The particular characteristics of the host spacecraft are determined by the size, pointing requirements, and data rates of the three SAMEX instruments, with the vector magnetograph having the greatest impact. Some general characteristics of the spacecraft for the purposes of this design study are listed as follows:

- (1) Spacecraft capacity for three instruments (solar vector magnetograph, H-alpha telescope, and soft x-EUV imager).
- (2) An approximate satellite weight in the 500 kg class.
- (3) Orbital placement in a near-Earth or Sun-synchronous orbit for a period of 3 years.

- (4) Telemetry rates compatible with TDRSS, with the magnetograph having typical rates of (a) 15-30 Kbps (daily average) at 80-100% of the time on a multiple access (MA) S-band channel of 50 Kbps for the total spacecraft, (b) 0.5 Mbps at 0-10% of the time on single access (SA) S-band, and (c) possible increases in data rates in excess of 3 Mbps with access to improved relay satellites.
- (5) Onboard magnetograph memory requirement of 1.5 Gb (equivalent to 10 vector magnetograms).
- (6) A volume of approximately $3 \times 1 \times 1 \text{ m}^3$ or less for each instrument.

In Table 67 the requirements on the host spacecraft for subsystem support are given.

Table 67. Subsystem support requirements.

	Vector Magnetograph	H-alpha Telescope	X-EUV Imager
Weight (kg)	150	50	100
Dimensions (l x d, m)	3.0 x 0.8	2.5 x 0.8	2.0 x 0.8
Power (W):			
Instrument	20	5	20
Thermal control	30	10	10
Average data rates (Kpbs)	30	3	17
Thermal range (°C)	0 - 15	0 - 15	0 - 15
Optical coalignment (between instrument and tracking system, in arc sec)	1	1	1
Pointing (all instruments)			
Angular range			
Line-of-sight		±0.55° from Sun center	
Roll		±180°	
Stability			
Line-of-sight		30 arc sec	
Roll		1.8 arc min for 5 minutes	
Stability with IMC			
Line-of-sight		0.25 arc sec for 5 minutes (225 enhancements)	
Accuracy			
Line-of-sight		1 arc sec	
Roll		2 arc min	
Orbital heliographic Doppler motion		50 m/sec	
Clock accuracy		1 sec	

The major constraints on the spacecraft are all imposed by the requirement for a spatial resolution of 0.50 arc sec. First, this means all components of each instrument must remain fixed relative to each other to within 25 microns, the scale of each pixel on the CCD chip (equivalent to 0.25 arc sec). This stability condition imposes very stringent limits on the allowable flexion induced into an instrument. Second, the high spatial resolution over the large field of view affects the telemetry requirements of the spacecraft. Third, the 0.5" spatial resolution requires commensurate accuracy in line-of-sight and roll pointing which impacts the requirements on the attitude and control system of the host spacecraft. Finally, the thermal design constraints on the mechanical housings and mountings of the optics will be considerable to provide the 0.5 arc sec resolution and the 1 arc sec coalignment accuracy. However, in the SAMEX magnetograph, the light is collimated between the collimator transfer lens at the primary and the camera focusing lens at the CCD; therefore, the system will be insensitive to line-of-sight changes in the path length. The long focal length of the camera lens also will reduce to some extent the sensitivity to variations in the mechanical rigidity of the spacecraft.

The roll of the spacecraft must be controlled to avoid image smearing. Assuming the spacecraft is pointed at the center of the solar disk, then a 0.25 arc sec pixel at the edge of the Sun subtends an angle given by $\arctan(0.25"/960") = 53$ arc sec. Hence the roll stability needs to be held within 2 arc min during observations for points near the limb to be resolved to 0.5 arc sec.

The error in pointing accuracy comes from three primary sources: bias error (e.g., star tracker error), long-term stability error (e.g., thermal drift), and short-term stability error (root-mean-square variation of pointing about the corrected mean, e.g., detector system noise). The bias error must be removed by in-orbit calibration; the other errors will be eliminated by the IMC.

The RCA's Defense Meteorological Satellite Program's (DMSP) BL-5D satellite has the subsystem capabilities to support SAMEX, except for pointing and stability where special sensors would be required. Modifications to the DMSP spacecraft would be needed for communications with the TDRSS. The capabilities of the BL-5D are listed in Table 68.

A list of other possible host spacecraft can be found in the 1984 report on the SAMSAT Pre-Phase A Study conducted by the MSFC Program Development Office.

Table 68. Subsystem capabilities of the DMSP BL-5D satellite.

Payload weight	2000 kg
Total average power	600 W
Data acquisition	60 Kbps
Pointing	12 arc sec
Stability	21 arc sec
Battery storage	50 AH (amp-hours)
Payload power	250 W
Communication rate	3 Mbps (S-band)
Communication service	AFSCF (Air Force system)
Payload interface	Integrated

5. Telemetry

Telemetry requirements of the observing modes and their compatibility with the TDRSS must be considered. For a 500-km, near-earth orbit, a TDRSS with two satellites could acquire data transfers for 92% of the time. For a full-Sun, Sun-synchronous orbit, the coverage of telemetry by TDRSS would be 100% of the time. The amount of data that can be handled depends on the TDRSS mode of operation. Table 69 lists the characteristics of the two modes of operations that are available with TDRSS (data are from the SAMSAT Study, MSFC, 1984). The multiple access (MA) mode utilizes a phase array which operates at an S-Band frequency and can support up to 20 users simultaneously. The maximum rate for the MA mode is 50 Kbps. The other mode is the single access (SA) mode which uses two high gain dish antennas. The SA mode can support high data rates of 3 Mbps at either S- or Ku-band frequencies. Normally a two-satellite TDRSS can handle only four SA users at any one time. The MA mode can support only two forward (command) links to any one spacecraft, which means that the command link must be shared with up to 20 users. Restrictions on the forward (command) and return (telemetry) links for the SA service mode are the same. These general restrictions on transmitting commands put a premium on mission planning and pre-planned observing sequences stored in the onboard computer software.

In the SAMEX magnetograph design, we have assumed a usual data rate of 50 Kbps for the entire spacecraft with 30 Kbps of this used by the magnetograph. This 30 Kbps rate for the magnetograph is assumed to be a continuous rate to handle the large amounts of data produced by the instrument. Special rates of 3 Mbps could be used to purge the mass storage or capture rapid events, but the timing of the use of these special rates would be critical since the total TRDSS requirements have to be considered and accommodated at all times (SAMEX will probably not be the only space-based system in orbit).

Table 69. TDRSS capabilities for a single satellite
(Ref. SAMSAT Pre-Phase A Study, MSFC, 1984).

	<u>Multiple Access (MA)</u>	<u>Single Access (SA)</u>
Maximum telemetry rate	50 Kbps	3 Mbps
Command availability	80-100%	10-20%
Telemetry availability	2 channels per 20 users	10-20%
Maximum average data rates	50 Kbps	600 Kbps
RF power (50 Kbps)	3.1 W	6.25 W (6.25% duty cycle)
Antenna size (50 Kbps)	1.2 m	1.2 m

The overall concept for the SAMEX data management must address several important elements in this general area of data transmission, including: (1) the total amount of data accumulated each day (~3-5 Gbits for average data rates of 30-50 Kbps); (2) the allowed telemetry rates (50 Kbps and 3 Mbps); and (3) the onboard software and hardware requirements for data, commanding, and engineering status of the instruments. All of these elements are driven by the magnetograph's data requirements since they are so much greater than those of the other two instruments. The SAMEX telemetry requirements are very similar to those of the Hubble Space Telescope. Consequently, if other space missions such as the Hubble Space Telescope, Space Station, and/or the Shuttle are in orbit at the time of the SAMEX mission, the overall mission planning will have to anticipate any possible saturation of the TDRSS system.

An estimate of the general telemetry rates for the magnetograph can be obtained by looking at three of the basic observational modes for the magnetograph: active-region, flare, and synoptic modes.

In the active-region mode, the field of view will cover a typical active region, and the instrument will monitor flare buildup and the associated magnetic changes with observations taken once every 20 minutes. A reasonable data rate can be achieved at the 20-minute cadence by resolving the field of view of 4.3' x 8.5' (256" x 512") with a 1 arc sec resolution that is achieved by binning the CCD pixels in a 2 x 2 mode. For a minimum of 5 images for each vector magnetogram with a word size of 14 bits, the data rate is

$$512 \times 1024 \times 14 \times 5 \text{ bits} / (20 \times 60 \text{ sec}) = 30.6 \text{ Kbps} ,$$

which can be handled by the TDRSS MA mode. On the other hand, at the highest possible spatial resolution of 0.5 arc sec and the

fastest cadence of 5 minutes, the data rate would be

$$1024 \times 2048 \times 14 \times 5 \text{ bits } / (5 \times 60 \text{ sec}) = 0.5 \text{ Mbps.}$$

This rate could not be transmitted by the TDRSS MA mode; the SA mode would have to be used. If the SA mode were not available, either the data rate would have to be reduced by onboard data compression schemes, or the data would have to be stored in the onboard mass storage unit and transmitted when the SA mode becomes available, or stored in a buffer storage unit and transmitted at a slower rate compatible with the TDRSS MA mode. This data rate, 0.5 Mbps, represents the upper limit for the data rates to be handled by the onboard computer system in any of the operational modes where the full 4.3' x 8.5' field of view is imaged with 0.5" spatial resolution.

To observe flare-associated changes, the flare mode must have the fastest cadence possible, a cadence determined by the temporal resolution of 5 minutes for a vector magnetogram. To keep the data rate within 30 Kbps and still achieve a spatial resolution of 0.5", we can limit the field of view to a smaller area encompassing only the (potential) flare site. The detailed magnetic changes around only the flare site will be monitored. This mode would utilize a buffer ring memory scheme and would be triggered by the H-alpha monitor to capture the flare activity as well as the preflare activity held in the memory ring. A fixed number, N, of magnetograms would be continuously updated in the memory in order to keep the last N magnetograms. The latest magnetogram would replace the earliest one. Then, at a flare onset as recognized by the H-alpha system, the preflare data would be retained for analysis. For a nominal field of view of 1 x 1 arc min (which can be changed to fit the specific area under observation), a pixel array of 240 x 240 will give the 0.5 arc sec resolution. The data rate then is

$$240 \times 240 \times 14 \times 5 \text{ bits } / (5 \times 60 \text{ sec}) = 13.4 \text{ Kbps ,}$$

which can be easily handled by the MA mode of TDRSS. In the event data would have to be stored over 1-hour interval at this rate (if, for example, no transmission were possible using TDRSS), 48.2 Mbits of memory would be required.

If only the line-of-sight component of the field is measured, only two images are needed and the cadence can be increased to the much faster rate of 1 magnetogram/minute. In this mode the data rate is

$$240 \times 240 \times 14 \times 2 \text{ bits } / (60 \text{ sec}) = 26.9 \text{ Kbps .}$$

The fastest rate is set by the temporal resolution of the line-of-sight mode, 38 seconds, but this produces rates of 42.4 Kbps, which would impose limitations on the amount of data obtained with the other two instruments.

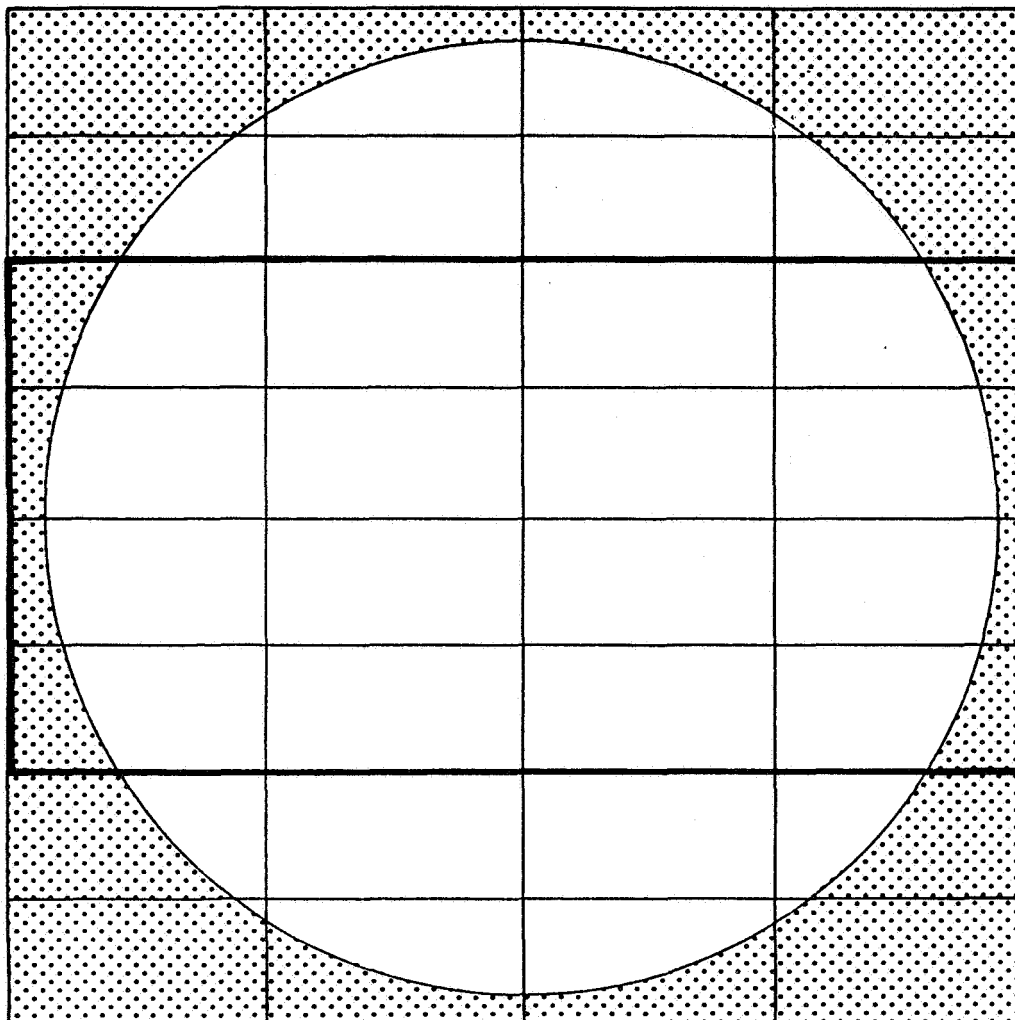


Figure 69. Mapping the solar disk with the SAMEX magnetograph. This figure indicates the number of images required to map the full solar disk with the SAMEX magnetograph which has a field of view of 4.3×8.5 arc min.

A synoptic survey mode will require a full disk image once a day; the full disk is imaged by mapping, a process requiring about 32 images utilizing the full 4.3' x 8.5' field of view as shown in Figure 69. To reduce data rates, the spatial resolution can be reduced to 2 arc sec by binning the CCD pixels in a 4 x 4 pixel-averaging mode. If onboard data processing discards all pixels that are outside the solar disk, then the number of 1" x 1" pixels needed to map the full Sun is given by the area of the Sun in arc sec²:

$$\pi (16 \times 60'')^2 = 2.9 \times 10^6 \text{ arc sec}^2.$$

The data rate is then given by

$$2.9 \times 10^6 \times 14 \times 5 \text{ bits}/(32 \times 5 \text{ minutes}) = 21.1 \text{ Kbps}.$$

A further data reduction of about 50% can be attained if only the active latitudes are mapped.

Data transmission by downlink telemetry must also include the housekeeping data: temperatures, encoder positions, voltages, time, roll rates, etc.

The preceding discussion shows that a 30 Kbps data rate for the SAMEX magnetograph will be adequate for most of the active-region, flare, and synoptic modes. Higher rates requiring the TDRSS SA mode will be generated if the full capabilities of the instrument are exercised. For data rates exceeding the 3 Mbps rate of the TDRSS SA mode, onboard storage and/or data compression techniques would be required. However, in the 1990's, improved telemetry systems may be available and could handle these large data rates.

IV. MISSION REQUIREMENTS

1. Introduction

In this section we look at some of the requirements that must be considered for the overall mission. These include satellite orbits, duration of the mission, and the ground-based facility that must be set up to provide mission control, data management, communications, and the facilities for scientific research with the SAMEX and ancillary data.

2. Orbital Considerations

Since the orbital placement of the SAMEX satellite affects the science to be gained from the magnetograph, as well as its engineering complexity and cost, the orbital options available must be considered. The options for orbital placement are listed in Table 70 for given altitudes and orbital inclinations.

Table 70. Options for orbital placement.

	Altitude (km)	Inclination (deg)
1. Low-Earth orbit:		
Low latitude	250-500	0-30
High latitude	250-400	30-60
Sun-synchronous and polar	200-500	60-97
2. Full-Sun, Sun-synchronous:	1390-3300	101-115
3. Geosynchronous:	41,000	0
4. L1-point:	1,700,000	23

As the orbital height and inclination increase, various parameters of the mission generally increase also; these include the size of the propulsion unit needed to deliver the satellite to orbit, the size, weight, and power of antennas and associated telemetry systems needed to return the data to ground, and the radiation levels encountered. Therefore, the lowest orbit compatible with the scientific goals would be the most attractive. Table 71 lists the minimum and maximum solar viewing time as a percent of an orbit, and the number of times and length of time that the Sun can be observed continuously during a year, for various orbital options. The time of orbital day can be determined by (Sherrill, 1981):

$$D = P \times \{1 - (\pi^{-1}) \arccos[(R^2 - R_e^2)^{1/2}/(R \cos B)]\},$$

where R is the orbital radius (km), R_e is the radius of the Earth (6371 km), and B is the angle between the Sun and the orbital plane of the spacecraft. The orbital period, P (min), is given by (Allen, 1973):

$$P^{2/3} = R/331.3 .$$

For 250- and 500-km orbits, the periods are 89.3 and 94.4 minutes, respectively. If the condition

$$R \cos(B) < (R^2 - R_e^2)^{1/2}$$

holds, then there will be continuous solar observations for the orbit. The beta angle is determined by the orbital inclination of the spacecraft and the solar inclination with respect to the Earth's equatorial plane. The solar inclination ϵ varies between $\pm 23.4^\circ$ (the obliquity of the ecliptic), depending on the time of the year. Hence, an orbital inclination of 57° can give continuous solar viewing during certain periods of time. The period of minimum continuous viewing of the Sun occurs when $B = 0^\circ$. For altitudes of 250 and 500 km, the length of the solar observing time per orbit is 52 and 59 minutes, respectively, about the duration of an average flare.

Table 71. Solar viewing time.

	Solar Viewing Time (% of orbit)		Continuous Viewing Time (number of hours/year)	
	Minimum	Maximum	Minimum	Maximum
1. Low-Earth orbit:				
Low latitude	59	66	0	0
High latitude	60	90	3	100
Sun-synchronous and polar	63	90	2	1100
2. Full-Sun,				
Sun-synchronous:	100	100	1	8760
3. Geosynchronous:	95	100	2	3500
4. L1-point:	100	100	1	8760

In Figure 70 we show three cases for the time of solar viewing (in percent): a geosynchronous orbit, a near-Earth polar orbit, and a near-Earth high-latitude orbit. The fluctuations in the percent of solar viewing for the high-latitude orbit are the result of the orbit nodal regression rate, $\approx 5^\circ$ per day, and the changes in the solar inclination.

One of the scientific goals of SAMEX is the observation of an active region for as long as possible as it traverses the solar disk, up to a period of 13 days or 312 hours. On the other hand, the minimum period of continuous solar observation must be at least several times the duration of a typical flare in order to capture the events surrounding its evolution to an unstable configuration; this means continuous observations of several hours. For a near-Earth orbit with a typical orbital period of 90 minutes, these conditions restrict the choice of orbit to a high-latitude one. For such an orbit, there would be two periods of ≈ 100 hours twice a year during which there would be continuous observations of the Sun. A near-Earth, Sun-synchronous orbit at an altitude of 500 km and with an inclination of 97.4° would provide continuous solar viewing for ≈ 265 days each year. The polar orbits would require launch of the SAMEX spacecraft from the Western Test Range.

The problems associated with the four basic orbital placements and the trade-offs involved must be considered. A geosynchronous orbit at 41,000 km and an orbit at the L1-point at 1,700,000 km are desirable in terms of the continuous or nearly continuous solar viewing (see Figure 70). The main objection to these higher orbits is the increased cost of putting a spacecraft into orbit. For a geosynchronous orbit, there is the additional problem of finding a parking orbit that does not interfere with other satellites. A geosynchronous orbit does remove the mission's dependence for telemetry on the use of relay satellites since a dedicated ground-based receiving system would provide direct access to the satellite. The L1 point at 0.1 AU is the point where the gravitational attraction of the Sun and Earth-Moon system is equal. A spacecraft inserted at the L1-point would require a propulsion system to maintain its orbit about this point. The additional distance would also require a larger antenna and a more powerful telemetry system.

A full-Sun, Sun-synchronous orbit can be obtained at an altitude between 1390 and 3300 km with orbital inclinations between 101.5° and 115.2° , respectively (SAMSAT Pre-Phase A Study, MSFC, 1984). A Sun-synchronous orbit is defined as an orbit which maintains a constant angle between the right ascending node and the right ascension of the Sun. An inclination of greater than 90° is required to provide the correct orbital precession rate. However, the cost of such a mission would be increased substantially because of the demands on the launch vehicle to place the satellite at this height.

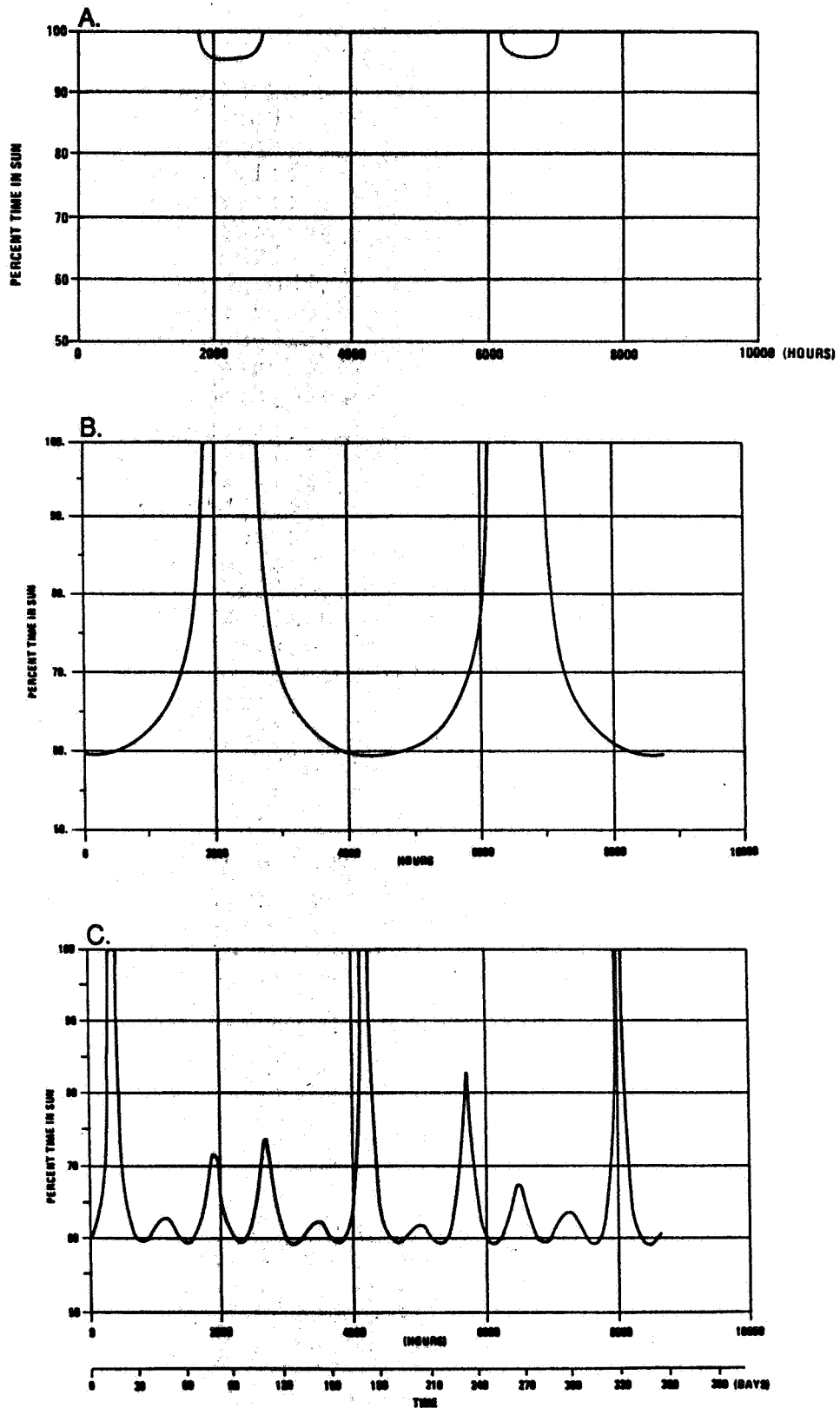


Figure 70. Periods of solar viewing for three different SAMEX orbits. Shown is the percent of time that the Sun can be observed over a period of 1 year. Top: geosynchronous orbit; center: near-Earth polar orbit; bottom: near-Earth high-latitude orbit.

Radiation levels must be considered since they pose significant problems for some orbits. For the geosynchronous orbit, radiation is due primarily to trapped electrons and protons, but the radiation levels are only a factor of 2 above the levels in a 500-km orbit. On the other hand, in the case of near-Earth orbits, decreasing a 500-km orbit by 100 km reduces the radiation level by a factor of 10. The flux of charged particles at the L1-point comes from the solar wind and cosmic rays, and is a factor of 10 or more lower than the geosynchronous orbit value. Radiation levels in a full-Sun, Sun-synchronous orbit are the maximum encountered, estimated at 10-100 rads per day for a spacecraft with 2 gm/cm² shielding.

A difficulty with low-Earth orbits, in addition to the problem of solar coverage, is telemetry. The option of direct telemetry to a ground station is not feasible, since a possible station like Thule Air Force Base (located at N76.5°, W68.7°), although passed over by a satellite in a near-Earth, Sun-synchronous orbit 76% of the time, would have actual contact for only 4 minutes per orbit. This would allow a dump of only 720 Mb of data at a 3 Mbps downlink telemetry rate, whereas the total amount of data accumulated each orbit by the magnetograph alone could be as high as 1200 Mb during 40 minutes of solar viewing per orbit: (150 Mb for each vector magnetogram)/(290 seconds per vector set) x 40 minutes. Hence, downlink telemetry cannot be achieved without the use of relay satellites for near-Earth orbits - the orbiting spacecraft would require the use of the TDRSS relay satellites. Since TDRSS usage would not be dedicated to SAMEX, its use by the Shuttle, Space Station, and other space missions could interfere seriously with the SAMEX telemetry.

The South Atlantic Anomaly (SAA) is the portion of the Earth's trapped radiation belt that comes closest to the Earth. The SAA is located over the South Atlantic at south 30° and extends over ±20° latitude at a level of 10 protons/cm²/sec for proton energies greater than 50 Mev. The electronic noise caused by electric charges in the SAA will affect SAMEX observations and other electronic equipment such as the sensors in the fine guidance system. The SAA will be encountered about 33% of the time each 24-hour period in a near-Earth, Sun-synchronous orbit; each encounter with the SAA lasts up to 20 minutes. These encounters will impact planning of the SAMEX experiment modes for near-Earth orbital missions.

The doppler shift, d , at a wavelength λ in a near-Earth orbit is given by

$$\frac{d\lambda}{\lambda} = v/c ,$$

where the maximum heliocentric velocity is given by the orbital velocity v at an orbital height h (km) using the relation

$$v = 2\pi \times (6371 + h)/P .$$

For 250- and 500-km orbits, the orbital velocities are 7.76 and 7.62 km/sec, respectively; at a wavelength of 5250 Å, these correspond to wavelength shifts of 136 and 133 mÅ, respectively. For the 250-km orbit, the spectral variation of the filter's passband about the heliocentric wavelength of the 5250 Å spectral line is given by

$$d\lambda = 136 \cos(Wt) \text{ mÅ},$$

where the orbital angular frequency W is 0.0012 rad/sec. Then the time rate of change of the passband shift $d\lambda$ is

$$d\lambda/dt = -136 \times W \times \sin(Wt) \text{ mÅ} = -0.163 \times \sin(Wt) \text{ mÅ/sec}.$$

Hence, as the SAMEX satellite approaches and recedes from the Sun, the magnetograph's spectral filter must undergo adjustments at a maximum rate of 0.16 mÅ/sec in a 250-km orbit. In comparison, the Doppler velocity due to solar rotation is 2 km/sec, leading to a wavelength shift of 35 mÅ at a wavelength of 5250 Å.

3. Duration of Mission

The SAMEX mission has as its main purpose the study of solar activity through an extensive research program using data from some very complex instruments. The observations carried out over the period of the mission are designed to provide images with high spatial and temporal resolution for extensive studies of active regions throughout their transit of the solar disk (13 days), for statistical studies of many active regions (for both basic research and the development of future monitoring programs), and for a partial look at the influence of the solar cycle. To provide the observations and data base required to carry out these goals, an extended 3-year mission should be planned. While shorter periods of observation with the SAMEX instruments would provide data of extreme value to research in solar physics, the full potential of SAMEX can only be realized over an extended mission. The arguments for an extended mission fall into two basic categories: the need to observe phenomena that last for long periods of time, and the need to develop a statistical basis for the study of certain events. Some phenomena having long time scales are:

- the solar cycle (11 years);
- active longitudes (months);
- coronal holes (months);
- prominences (months);
- active regions (weeks to months).

Obviously, a mission lasting only a week would not provide sufficient observational coverage of these events.

The number of active regions observed in a 3-year mission would be sufficient to conduct an extensive program of active region research. In Figure 71 (Wilson et al., 1987), the 12-month running average of the number of sunspot groups for the last few cycles is presented. These numbers were derived from the Astronomical Observatory of Rome Report on Solar Phenomena. The monthly averages mostly lie within the envelope of ± 2 groups. The average number of sunspot groups ranges from about 8 groups on the disk at any one time to 1 group per month near solar minimum. Near solar maximum, there will be an average of ~ 5 groups daily that can be observed. During a year's period with an average sunspot number of ~ 100 , there are ~ 500 new groups on the Sun (Allen, 1973). These numbers indicate there will be sufficient activity in a 3-year mission to provide good statistics to study the various magnetic configurations that occur in active regions.

4. Ground Operations and Data Management

4.1 Control Center

A full-time operations facility will be required for the SAMEX mission to provide data management and archiving, data analysis and display, and the overall direction for the SAMEX scientific program. The details of such an extensive and complex facility are not completely defined herein, but a general concept of the basic requirements for this facility is given. To better define these requirements, we have referred to the ground-based operations and facilities that have been established for a number of orbiting observatories in the past; these include OAO-2 (1968), OAO-3 (1972), Skylab-ATM (1973), HEAO (1977), IUE (1977), Solar Maximum Mission (1980), Spacelab II (1985), and the Hubble Space Telescope (to be launched). In common with these space observatories, the SAMEX ground-based control facility would include a Payload Operations Control Center (POCC), Science Operations Center (SOC), Mission Operations Center (MOC), and Data Operations and Management Center (DOMC). The facility would be regarded as a national solar facility, and it is a critical requirement for a productive SAMEX program. A breakdown of these operational centers is shown in Table 72 along with some of their functions.

A primary objective of this ground control facility is to carry out the scientific goals of SAMEX in the most efficient and comprehensive way by providing a facility where the appropriate observational programs are planned on the basis of information from real-time analyses of the SAMEX data. The amount of data that the SAMEX magnetograph will generate will be extensive. Thus the handling and analysis of these data in real time (or nearly real time) will be prime functions of the facility and primary drivers in its design.

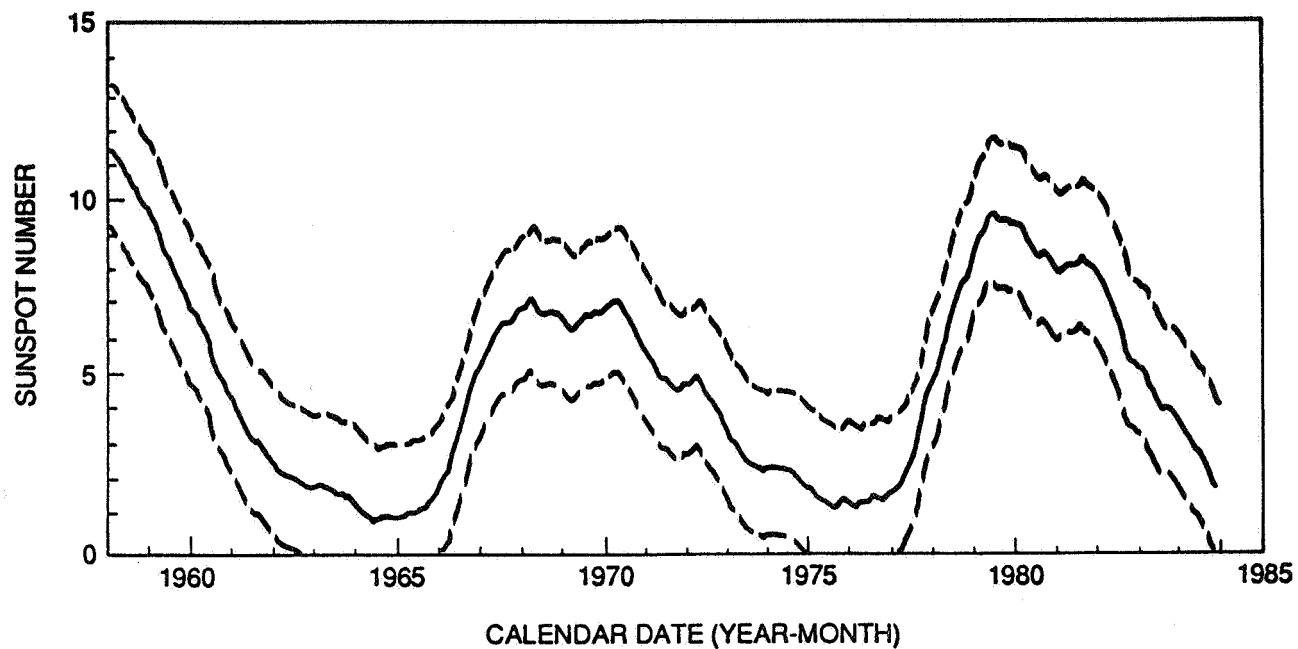


Figure 71. Number of sunspot groups seen during the past several solar cycles. Shown is the 12-month running average of the number of sunspot groups calculated from 1958 to 1985.

Table 72. Operational control facility.

-
1. Mission Operations Center (MOC)
 - Overall operation of the spacecraft
 - Project management
 - Position and attitude of spacecraft
 - Target acquisition
 - Troubleshoot spacecraft failures and anomalies
 - Command management
 - Spacecraft safety and health
 2. Science Operations Center (SOC)
 - Research operations
 - Timeline development for integrated observing schedules
 - Investigators Working Group interface
 - a. Science program guidelines
 - b. Selection of science programs
 - c. Selection of and interaction with guest investigators
 - Scientific staffing
 - Development of science observations, procedures, and capabilities
 - Scientific information and publicity
 - Scientific performance evaluation
 3. Payload Operations Control Center (POCC)
 - Operation of individual payload instruments
 - Quick-look assessment
 - Technical support
 - Software analysis
 - Calibration
 - Command preparation
 4. Data Operations and Management Center (DOMC)
 - Telemetry
 - Data archiving and temporary storage
 - Dissemination of data and observing plans
 - Scientific data processing and analysis
 - Data communication interface
 - Computer systems and terminals
 - Data acquisition, formatting, and special data processing
 - Command decoding, verification, and storage
 - Command processor
 - Spacecraft software generation
 - Ground-based software generation
 - Batch processing
 - Ancillary data reception and processing
 - Data base management
 - Ephemeris data generation
-

In the Data Operations and Management Center, the data will be formatted for processing, dispersal, and archiving. Here the engineering data from the spacecraft and instruments will be extracted from the data stream for real-time analysis and immediate display, and the scientific data will be processed for calibration, analysis, and archiving.

The data from an instrument must flow through an elaborate path to reach the final data storage system: from the instrument to the spacecraft's telemetry system, to the relay satellite(s), to ground receiving stations, to the ground control facility, through the processing computers, to temporary data storage, and finally to the mass storage archiving system. Because of the complexity of this network, the data stream must be monitored continuously to insure that no data are lost. All the data must flow through a standard processing system which might include decommutation, synchronization, and bit error corrections. It is therefore extremely important that the system is accurately designed to be compatible with the data rates produced by the SAMEX magnetograph, averaging almost 3,000 Mbits per day of image data.

Software development will play a major role in balancing the use of onboard processing against that of ground-based processing. The cost for real-time processing is greater than for batch processing, and an analysis of this trade-off for the magnetograph operation will have to be considered.

Telemetry of data through the TDRSS network will have to be carefully scheduled and monitored to avoid data losses either because of onboard memory overflows or during the telemetry process (the high-gain antennas must be pointed at the TDRSS during all transmissions). Because of the constrained capabilities of the spacecraft's telemetry, there will have to be a continuous, well-conceived interweaving of data flows from observations with the individual instruments into the onboard data handling system.

Large data arrays of 1024 x 2048 pixels, each with a dynamic range of 14 bits, and the large data rates produced will both require a special data processing facility. This facility will have to incorporate equipment and techniques for interactive analyses of photographic images and vector magnetograms, and real-time analyses of the temporal evolution of vector magnetic fields and how it correlates with photospheric flows and photospheric, chromospheric, and coronal imagery. Display devices, software, hardware, and array processors will be crucial elements of the system configuration and data manipulation.

The basic requirements for the facility include: office space for scientists, operational, and support personnel, and staff; data analysis facilities; work areas; a photographic laboratory; conference rooms; computing and data analysis equipment; and libraries.

4.2 Computers

The computers for the ground control facility must be able to process SAMEX scientific data at the rate of 50 Kbps on a daily average. The magnetograph data, coming in at an average daily rate of 30 Kbps, will be in the form of polarization images; processing and calibration of pairs of these images will be required to transform the polarized intensities into images of the photospheric vector magnetic field. Also, the computers must handle the processing of many magnetograph images simultaneously to generate displays and to perform image registrations, transformations, etc. Special processing of the data to perform magnetostatic extrapolations, and analyses for temporal and spatial variations will require array processors, large virtual memory systems, and sophisticated graphic display devices.

Much of the processing will be at digital video rates, in order to bring up and manipulate the images in real time. Hardware devices and software programs are now being developed for large data array processing. Many of the near real-time analyses can be performed with VAX-class computers and the 64-bit, 64-Mb array processors currently available.

4.3 Mass Storage

A mass storage facility will be needed to handle the data generated at an average rate of 50 Kbps by all three of the SAMEX instruments. If we assume a 3-year lifetime for the SAMEX mission, the facility will require a mass storage device with a capacity of 4.73×10^{12} bits:

$$(3 \text{ years})(365 \text{ days/year})(86400 \text{ sec/day})(50 \text{ Kbps}).$$

An archival storage and retrieval system with this capability is being developed at MSFC as part of the NASA Data System Technology Program and the Space Plasma Analysis Network (SPAN). The SPAN system uses 125 RCA optical disk readers on fiber optics buses and an argon gas laser with a spot size of 0.5 micron to read high density disks. The online capacity of this system is 9.75×10^{12} bits with data rates of 50 Mbps. The storage requirement for the optical disks is only 52 square feet; storage of an equivalent amount of data on magnetic tapes would require 8,250 6250-bpi tapes occupying about 796 square feet. Also, the cost of using optical disks is a factor of 10 less than the cost for magnetic tapes.

Another proposed system, the Harris Corporation's "fifth generation systems" would provide a 10^{15} bit Mass Archival Storage and Retrieval (MASTAR) Optical Memory System (Reddersen and Ralston, 1980). The development of large online data storage is an active technology in today's computer world which assures the availability of adequate mass storage for the SAMEX mission. In fact, storage of the entire data set from a 3-year SAMEX mission is feasible and would provide an interactive data

base for the solar and space physics communities. Various companies (e.g. 3M, Kodak, Sony, RCA, and Hataish) are developing giga-byte, high-density optical and electromagnetic recording devices. The Global Oscillation Network Group (GONG) project has reported an audio device under development, the Rotary Digital Audio Recording Technology (R-DAT), with a cassette holding 1.38 Gbytes in a $7.3 \times 5.4 \times 1.05 \text{ cm}^3$ container (Harvey, 1987).

Optical disk technology will also be used to store ancillary data received from ground-based observatories during the mission, and to facilitate data exchanges. This will encourage and enhance active participation in the SAMEX scientific program by the solar community and interested agencies.

4.4 Data Analysis Facility

Hardware for a data analysis system will include a computer system and graphics work stations. The computer requirements for image handling, data processing, numerical simulations, and computer modeling were discussed previously. The hardware for image processing is considered here and is based in part on the National Oceanic and Atmospheric Administration (NOAA) Space Environment Laboratory's (SEL) Solar Imaging System (SELSIS) and Data Analysis and Display System (SELDADS), and on requirements established from operations with the MSFC vector magnetograph.

The graphics work stations require hardware for accessing, receiving, and analyzing various solar images and magnetograms. They must also be able to perform analyses involving computer interactions with the data archival system and the mainframe computer, and be able to interact with the computers at the network observatories. The individual stations should have the capabilities to provide a wide variety of image displays and perform analyses such as image manipulation, enhancement, horizontal motion studies, comparisons (arithmetic, etc.), overlays, quantitative analysis of results, and subjective (interactive) analysis. In addition, they should incorporate a range of techniques in image analysis that is comprehensive enough to easily accommodate newly developed software. The use of color graphics in displaying the vector magnetic field has proven very effective in the analysis of data at the MSFC vector magnetograph facility and should be considered for the graphics work stations. To realize the full potential of the SAMEX data - data describing the photospheric vector magnetic field, plasma motions, and the morphology and physical properties of the magnetically-controlled plasmas in the chromosphere and corona - will require a highly developed technology in image processing and analysis.

Other solar, interplanetary, and near-Earth space data should also be made available for interactive analysis. This data base would include soft x-ray emission levels, solar event information, and geophysical parameters. These data will be important for researchers interested in the SAMEX data from the

aspect of developing techniques for predicting solar activity. The major hardware requirements for the data analysis facility are listed in Table 73.

Table 73. Computer facilities for the data analysis system.

-
1. Data Analysis Computer System
 - DEC VAX-8300 computer system
 - Four DEC RA81 disc memories (500 Mb each)
 - CSP/MAP-6420 array processor for image manipulation
 - DEC HCS 50 Star Coupler cluster component for computer clustering
 - Optimem 1000 laser disc WORM units with 1 Gb storage at 5 Mbps transfer
 - Chromatics CX-1260 high resolution graphics terminal
 - Matrix camera
 - Line printers (LA50)
 - Three IBM PC/AT terminals and DEC terminal software emulator
 - Flat bed plotter
 - ISSCO Disspla graphics software
 - Mathematical software package
 2. Individual Graphics Work Stations
 - Four Apollo analysis stations
 - Three high resolution display terminals
 - Optimem 1000/S laser disc reader at 1 Gb storage per disk
 - Line printers and plotters
 - Two modems (9600 baud)
-

The solar community worldwide should be strongly encouraged to carry out cooperative and supportive ground-based observations in the optical and radio wavelengths to enhance the scientific returns from SAMEX. A data base for these observations should be incorporated in this data analysis facility. The GOES Advanced Payloads AP-4 and AP-5 will provide additional data to enhance the SAMEX data, e.g., full-disk x-ray data. A network linking these various data bases should be established to facilitate carrying out the SAMEX research programs and enable solar research with these data to continue over an extended period of time well after the completion of the SAMEX mission. This continuation will assure that full use is made of this unique data base.

4.5 Communications

The SAMEX data analysis facility will be one of the major centers for solar science throughout the world. Hence it will be necessary to provide communication facilities for receiving data from observatories around the world and transmitting data to them. The solar community would participate in the data analysis, and contribute to the SAMEX mission through joint observing programs. This collaboration with the solar community could be carried out most effectively and efficiently using a computer system network designed to communicate both image data and written text. A model for such a system has already been established - the Space Plasma Analysis Network (SPAN) currently used by the solar-terrestrial and interplanetary physics communities to analyze space data. The SPAN facility provides electronic mail, document browsing, access to distributed data bases, facilities for numeric and graphic data transfer, access to Class VI computers, and entry to gateways for other networks (Green and Peters, 1985). The network uses existing Digital Equipment Corporation computers as network nodes and communicates over leased lines using DECnet protocol. A similar network for the SAMEX mission would address the requirements for scientific data exchange and collaborations which will be necessary for the ultimate success of SAMEX.

V. DATA REDUCTION AND ANALYSIS

1. Introduction

The reduction and analysis of the SAMEX data will be a major effort in the SAMEX project. To provide ample time for a full range of synoptic observations and to utilize completely the full potential of the SAMEX instruments, we must think in terms of a 3-year mission. Such an extended observational period will produce an extensive set of data. And these data must be converted into formats that can be used easily and conveniently by the whole solar community. Thus the processes to handle the giga-bits of data from SAMEX and the software to convert these bits into usable information must be defined and developed. Once converted, the data can be used in many different analyses to derive physical parameters important in our understanding of solar activity and to model the three-dimensional characteristics of magnetic fields in active regions. In this section we outline some of the analyses we anticipate will be used with the data from the SAMEX vector magnetograph. Much of the material presented is based on operational techniques that have been developed for the analysis of data from the MSFC vector magnetograph.

A general description of the data analysis facility for SAMEX was given in section IV.4.4. To be effective in carrying out most of the studies in solar activity, this research facility must have rapid access to the data as they are received at the ground receiving station for the mission. This requirement for data reception in nearly real time arises from the need to monitor carefully the performance of the SAMEX instruments and to issue new or updated commands to the SAMEX satellite based on analysis of the transmitted data. However, the complete analysis of the data will be an extensive project which cannot be done online. For example, the study of the growth and evolution of an active region will extend over a period of many days. For these and other studies, only a selected set of data will have to be sent in nearly real time to the SAMEX research facility. Most of the data can be stored on high-density optical disks at the receiving station and thus reduce the cost of re-transmitting data to the SAMEX data analysis facility. However, the data must be received in a timely manner at the SAMEX center (for example, by using an overnight delivery process). The extensive software needed for analysis of the real-time data as well as the data received at later times is outlined in the following sections.

2. Software for Data Analysis

Existing software packages can be used for certain parts of the data reduction and analysis. There are various integrated software packages designed specifically for scientific data that provide sophisticated data processing routines, high-quality

graphics programs, and image processing and display. One such package, the Image Reduction and Analysis Facility (IRAF) developed at the National Optical Astronomical Observatories (NOAO), provides general reduction and analysis programs for use with astronomical data. The Science Data Analysis Software (SDAS) system developed for analysis of data from the Hubble Space Telescope (HST) provides a set of application programs for specific astronomical investigations related to the HST. Many of the programs from both of these software packages can be adapted for use with the SAMEX data. In fact, they already form an integrated software package to a certain extent since the SDAS applications software runs under the control of IRAF command language. The SDAS package uses four external software libraries: (1) the scientific graphics display system from NCAR (National Center for Atmospheric Research, Boulder, Colorado), (2) the mathematical routines and numerical analysis software from NAG (Numerical Algorithms Group, Oxford, U.K.), (3) the mathematical and statistical libraries from IMSL, Inc. (Houston, Texas), and (4) the IP8500 image display library for the Gould-DeAnza image display device. Additional, specific software will have to be developed to interface these packages with device-dependent software and with I/O software for retrieval and display of the SAMEX data. The features that the system would then have include mathematical and statistical analyses, data interpolation, filtering, curve fitting, data editing, interactive graphics, image processing and display, array processing, two-dimensional transforms, and modeling codes.

These software packages can be used as an initial foundation from which to build the software package of the SAMEX system in a VAX environment; more specialized software can be added to this core to address problems in analysis that are specifically related to the data from the SAMEX instruments and the experiments carried out with these instruments. The software developed over the past decade for reduction and analysis of data from the MSFC vector magnetograph will provide some of the specialized programs for analysis of the SAMEX magnetograph data. In Table 74 there are listed some of the MSFC data reduction programs to correct the observational data for cross talk and biases, to transform the measured polarized intensities into field components, and to display the field components in various formats.

Some auxiliary programs have also been developed to aid in the reduction and analysis of MSFC vector magnetograms; these are listed in Table 75.

Many computer programs have been developed at MSFC to perform some specific analyses once the calibrated vector magnetic field has been derived from the raw data. Some of these are listed in Table 76.

Table 74. MSFC computer programs for data reduction.

-
1. Calculation and filtering of I, Q, U, V
 2. Photometric calibration of I
 3. Correlation plots for cross talk
 4. Distribution plots for biases
 5. Absolute calibration
 6. Calculation of calibrated magnetic field components (B_L , B_T , $|B|$, ϕ)
 7. Resolution of 180° ambiguity in ϕ :
 - comparison with potential field
 - dB_z/dz method (Harvey, 1969)
 - nearest neighbor method
 - interactive method
 8. Selection of contour levels and plot formats
 9. Contour program to contour B_L , B_T , $|B|$, J_z , α , etc.
 10. Plot contours
 11. Plot azimuth with (without) arrows
 12. Plot three-dimensional surface
-

Table 75. Auxiliary programs.

-
1. Filter data
 2. Label plots
 3. Draw grid on plot
 4. Change contour levels
 5. Format data for printing
 6. Subtract two files
 7. Average two files
 8. Divide two files
 9. Add constant to file
-

Table 76. Programs for derived physical parameters.

-
1. Calculate J_z
 2. Calculate $\alpha = \mu_0 J_z / B_z$
 3. Transform vector into heliographic coordinate system
 4. Locate neutral line(s)
 5. Plot any quantity along neutral line
 6. Calculate and plot distribution of angular shear
 7. Calculate and plot source field B^*
 8. Calculate and plot dB_z/dz (from $\nabla \cdot B = 0$)
 9. Calculate magnetic energy (Low, 1982)
 10. Calculate Lorentz force and torque
-

Finally, more involved analysis programs have been developed at MSFC to treat the problems associated with radiative transfer of polarized light and to model the extension of the photospheric field into the transition region and corona; these programs are listed in Table 77. More detailed developments of these analyses are presented in the following sections and in Appendix E.

Table 77. Radiative transfer and modeling programs.

-
1. Program to calculate I, Q, U, V as a function of λ with (without) magneto-optical effects using
 - (a) Unno theory (1956)
 - (b) Kjeldseth-Moe theory (1968)
 2. Program to calculate I, Q, U, V passed through a filter as a function of λ with (without) magneto-optical effects using Unno and Kjeldseth-Moe theories
 3. Program to calculate percent linearly and circularly polarized light passed through a filter as a function of magnetic field strength with (without) magneto-optical effects using Unno and Kjeldseth-Moe theories
 4. Program to calculate transmission profiles for Fabry-Perot and birefringent filters
 5. Program to calculate magnetic field distribution from an arcade of current loops
 6. Programs to calculate potential field using B_L data
 7. Programs to calculate linear force-free fields using B_L data
 8. Programs to calculate non-linear force-free fields using B_L and B_T data
-

With this software developed by MSFC and the software packages IRAF and SDAS, the SAMEX data analysis effort will have an excellent foundation on which to develop the data analysis programs for the SAMEX experiments.

3. Theoretical Modeling

The computer programs for theoretical modeling can be divided into two main areas: magnetostatic (MHS) and magnetohydrodynamic (MHD) codes. Computer algorithms for both MHS and MHD models have been developed for the analysis of MSFC vector magnetograph data, and, supplemented by images of solar features observed in the chromosphere, transition region and corona, these models have been used to investigate the morphology and configuration of the magnetic field in active regions. The SAMEX data will provide coaligned and cotemporal images and magnetograms of active regions which will be ideal for use in these modeling studies.

3.1 Magnetostatic Codes

A major effort in the analysis of the SAMEX magnetograms would be directed toward using magnetostatic codes to determine the deviations of the observed fields from configurations expected if the fields were either potential or force free. It is now clear that the structure of an active region is determined largely, if not entirely, by the structure of its magnetic field, and this field topology is itself determined principally by the distribution of magnetic flux in the photosphere (cf. Stenflo, 1983; Priest, 1982; Orrall, 1981). There exist well-established methods for measuring this magnetic field in the photosphere. As indicated by Harvey (1977), these methods are based on measurements of the Zeeman effect in photospheric spectral lines. While lines formed at different heights in the lower solar atmosphere have been used to infer the three-dimensional structure of magnetic fields higher in the atmosphere (Giovannelli and Jones, 1982), the range of heights is generally quite small (~ 1000 km) compared to the height reached by magnetic fields above active regions as inferred, for example, from x-ray images (Eddy, 1979; Berton and Sakurai, 1985). High resolution microwave interferometry can be used to observe magnetic structures high above the photosphere (i.e., in the corona), but, to interpret such observations in terms of coronal magnetic field strengths and topology, it is still necessary to assume some model for the coronal field and plasma (cf. Holman and Kundu, 1985). Consequently, to infer the large-scale, three-dimensional structure of the field B , we must use some computational model to extrapolate the magnetic field distribution above the photosphere, but this requires us to make some assumptions about the currents J flowing through and above the photosphere, since these currents are the source of any non-potential magnetic fields in the atmosphere. A particular assumption about the currents leads to the magnetostatic problem of determining the magnetic field as a function of (1) the observed boundary conditions on some surface (the photosphere) and (2) the assumed distribution of electric currents. Mathematical methods are then developed to solve this particular mhs problem, and three-dimensional magnetic field lines are constructed and compared with known field line "tracers" such as the fibrils and loops observed in the chromosphere (in H-alpha images) and corona (in soft x-ray images) of active regions. Excellent summaries are given by Levine (1975, 1976) of methods to extrapolate solar magnetic fields from photospheric magnetograms and prescriptions for comparing such models with observations.

If all of the currents giving rise to the magnetic field are subphotospheric, the field in the photosphere and overlying atmosphere is a potential field; this field configuration represents the state of minimum magnetic energy and its three-dimensional structure is uniquely determined by the photospheric distribution of either B_L , the line-of-sight component, or B_T , the transverse component (Schmidt, 1964; Sakurai, 1982; Hannakam et al., 1984). The equations for the case of a potential field are derived from the two Maxwell equations

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{J}$$

and

$$\nabla \cdot \mathbf{B} = 0.$$

If $\mathbf{J} = 0$, the field \mathbf{B} can be derived from a scalar potential ϕ , and this condition leads to the potential equation

$$\nabla^2 \phi = 0,$$

which has a unique solution determined by the photospheric boundary conditions (B_L or B_T). Once this unique solution is derived, the field \mathbf{B} is known everywhere and hence field lines can be calculated from any given starting point, usually on the photospheric surface. In Figure 72 there is displayed a set of field lines obtained from the Schmidt method of calculating a potential field using a line-of-sight magnetogram from the MSFC magnetograph.

If there are electric currents in or above the photosphere, and if these currents are everywhere parallel to the magnetic field, the configuration is called "force free," since the Lorentz force given by $\mathbf{J} \times \mathbf{B}$ will be zero. In this case, the Maxwell equation relating current density \mathbf{J} to the field \mathbf{B} reduces to

$$\nabla \times \mathbf{B} = \alpha \mathbf{B},$$

where α is a scalar function that in general varies in space and time. If α is assumed to be a constant value independent of the spatial coordinates, then once again a unique solution for the field can be found in terms of the photospheric boundary condition B_L or B_T , so that specifying either B_L or B_T on the photosphere determines the three-dimensional structure of the magnetic field for a given alpha, where alpha is below a critical value α_{\max} (Nakagawa and Raadu, 1972; Chiu and Hilton, 1977; Alissandrakis, 1981). The existence of a maximum value of alpha (α_{\max}) comes from the mathematical treatment of extrapolating the magnetic field using only the line-of-sight component of the field. As α increases, the magnetic field becomes more and more twisted, thus creating stronger currents. The fields from these currents force the configuration to open up and reach a point where the assumed exponential decay with height becomes invalid. The field arising from the currents must go into changing the transverse vector field at the photosphere while the longitudinal photospheric field is given (i.e., fixed) in the constant-alpha models. These computed force-free transverse fields can then be compared with the observed transverse field using data from a vector magnetograph; the degree to which the observed and computed photospheric fields match will give an indication of the validity of the basic constant-alpha, force-free field assumption.

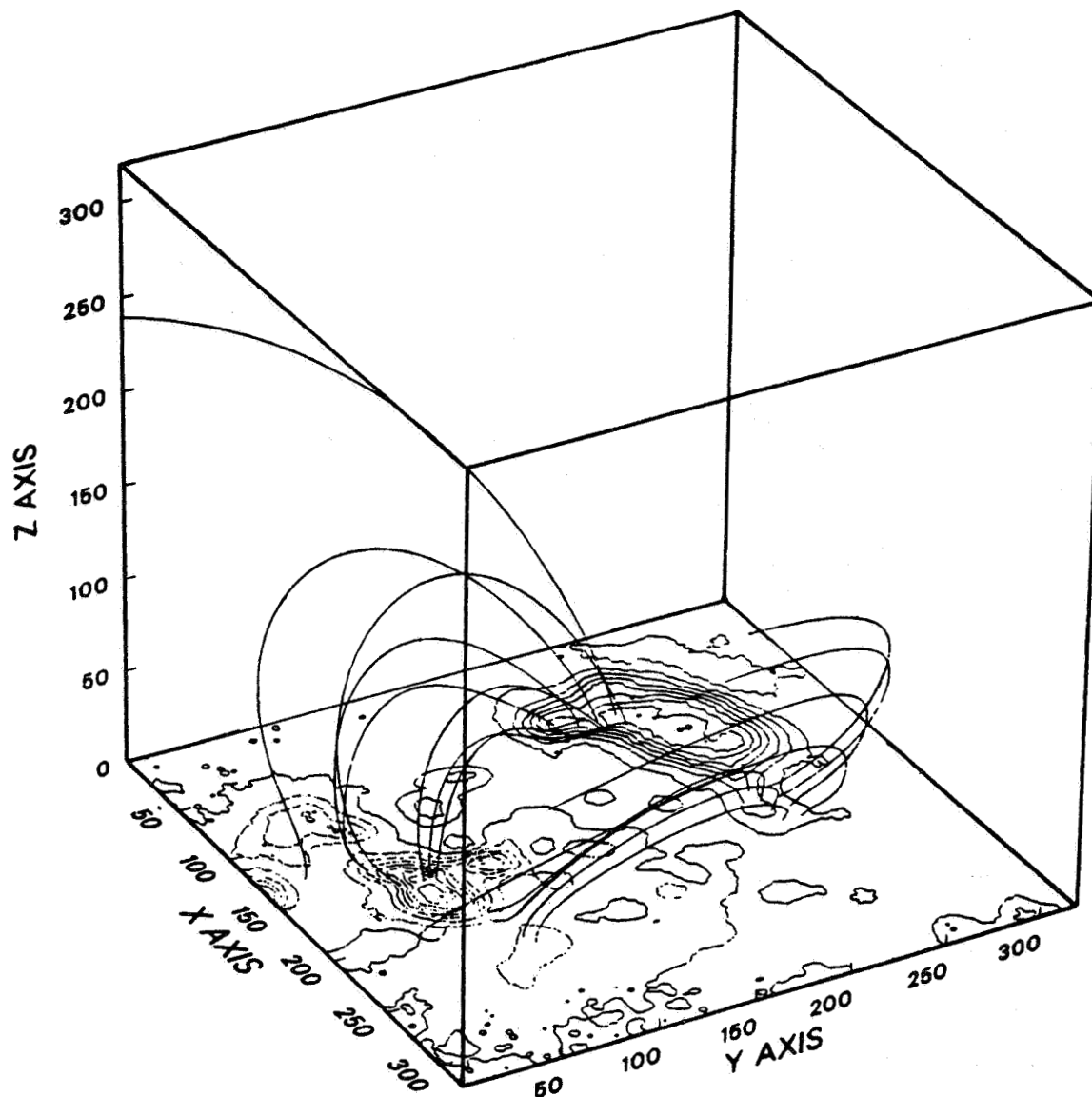


Figure 72. Computed potential field of an active region observed with the MSFC vector magnetograph. The Schmidt method was used to compute the potential field distribution for an active region observed on 23 September 1980. The contours delineate the line-of-sight component of the photospheric magnetic field measured with the MSFC magnetograph. This component was then used as the boundary value for the Schmidt calculation to derive the field lines shown. The scales represent distances in arc seconds.

For the constant-alpha field, the methods developed by Welck and Nakagawa (1983) and Alissandrakis (1981) for force-free magnetic field computations can be used; for the case $\alpha = 0$, the potential (current-free) case is recovered.

In all likelihood, α is not a constant everywhere. (Indeed, there may be some currents not parallel to the magnetic field, but this situation would require compensating forces to locally balance the Lorentz force $\mathbf{J} \times \mathbf{B}$ in order to maintain a stationary configuration.) Methods are now being developed to calculate non-constant-alpha solutions (Sakurai, 1979, 1981; Pridmore-Brown, 1981; Wu et al., 1987b). At this time these methods are not yet developed to the point of being proven analytical techniques. However, these and other methods should be fully developed and available for use in analyzing the SAMEx data.

Despite their limitations, potential and constant-alpha, force-free field models still are useful analytical tools. From the line-of-sight component of the observed vector field, three-dimensional fields can be computed and compared with tracers of the field above the photosphere; such comparisons indicate the extent to which the actual solar field can be approximated by the chosen model. In addition, the model's transverse field can be computed in the photosphere and compared directly with the observed transverse field. Using potential and force-free models for these comparisons will show where in an active region the observed photospheric field is non-potential and how much it deviates from a potential configuration.

Alternatively, the observed transverse field can be analyzed directly for the presence of currents entering or leaving the photosphere (Krall et al., 1982; Haisch et al., 1986; deLoach et al., 1984; Hagyard et al., 1985). These currents are calculated from the expression

$$\mu_0 \mathbf{J} = \nabla \times \mathbf{B}.$$

Since the z-component of the curl can be computed from the observed transverse field, we can calculate the distribution of the z-component of the electric current density over an active region. The technique developed by Hagyard et al. (1981) provides further information on the currents flowing above the photosphere. In this method the source field \mathbf{B} is derived from the difference between the observed field and the potential field, and then the source field is analyzed to infer the configuration of currents that produced it. One can also calculate an average vertical current density over the entire active region from the expression

$$\mu_0 \langle J_z \rangle = \alpha \langle B_z \rangle,$$

and find the total integrated current over the area observed.

The non-potential characteristics of the magnetic field above the photosphere of a particular active region can also be determined by comparing observed chromospheric field tracers with potential and linear force-free fields computed from the observed line-of-sight field. In Figure 73 a comparison of the potential and observed transverse field is shown. These comparisons allow us to visualize how the stress in the magnetic field changes with height above an active region.

Under the assumption that the magnetic field of an active region is a force-free field, we can derive an integral equation that yields the magnetic energy E_M of the field and the degree to which the field deviates from the force-free condition (Aly, 1984; Low, 1982; Molodensky, 1974; Stratton, 1941). If we consider a general magnetic field \mathbf{B} in the half plane above the photosphere with the lower boundary formed by the photosphere, then the net force \mathbf{F} exerted on a volume distribution of current \mathbf{J} in the field \mathbf{B} is

$$\mathbf{F} = \int_V \mathbf{J} \times \mathbf{B} \, dV,$$

where V is the volume defined by a closed surface S with its lower boundary A on the photospheric surface. Since

$$\mathbf{J} = \frac{1}{\mu_0} \nabla \times \mathbf{B},$$

the net Lorentz force is given by

$$\mathbf{F} = \frac{1}{\mu_0} \int_V (\nabla \times \mathbf{B}) \times \mathbf{B} \, dV. \quad (1)$$

One can show that this may be written

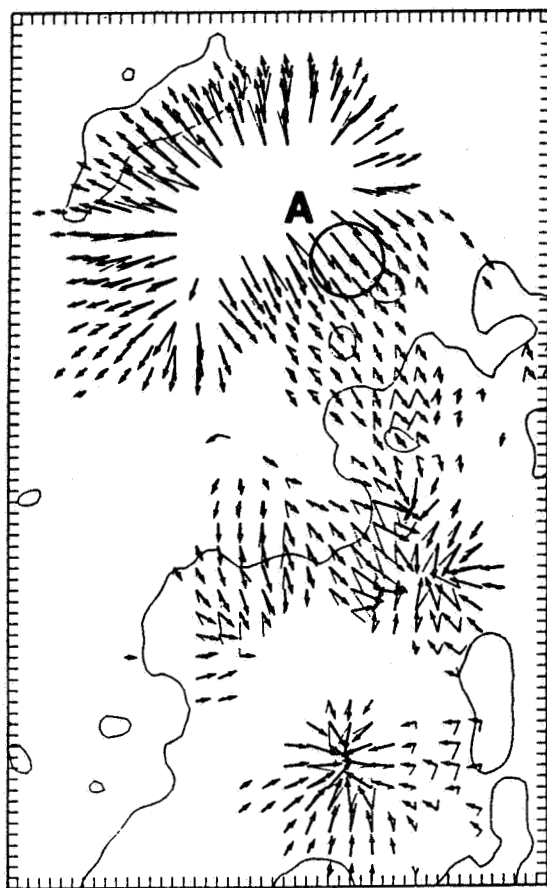
$$\mathbf{F} = \frac{1}{\mu_0} \int_S \mathbf{B} \mathbf{B} \cdot d\mathbf{S} - \frac{1}{2\mu_0} \int_S B^2 \, d\mathbf{S}.$$

If we assume the field \mathbf{B} is sufficiently small over all of S except on the photospheric area A (the x, y plane), then we have these equations for the three components of the Lorentz force:

$$F_x = - \frac{1}{\mu_0} \int_A \int B_x B_z \, dx \, dy, \quad (2)$$

$$F_y = - \frac{1}{\mu_0} \int_A \int B_y B_z \, dx \, dy, \quad (3)$$

$$F_z = - \frac{1}{2\mu_0} \int_A \int [B_z^2 - (B_x^2 + B_y^2)] \, dx \, dy. \quad (4)$$



(a)



(b)

Figure 73. Non-potential characteristics of an active region observed with the MSFC vector magnetograph. Magnetograms of the line-of-sight and transverse components of the photospheric magnetic field observed in active region 2684 on 23 September 1980 are used to define areas where the magnetic field deviates from a potential configuration. These areas can be identified by comparing the observed transverse field with the transverse component of that potential field having the observed line-of-sight field as its lower boundary. The field of view in both panels is 125×200 arc sec. (a) Comparison of the computed potential transverse field with the observed component. The magnitude and direction of the transverse component of the magnetic field are indicated by the length and orientation of the line segments; the potential field is designated by the line segments with arrowheads. The observed field is seen to be only mildly non-potential. (b) Potential field lines projected onto the photosphere. The field lines shown in Figure 72 are seen in projection at the photosphere where the potential transverse field is plotted as the directed line segments. The field line projections are seen to be aligned fairly well with the photospheric field.

If the field is force-free, $F_x = F_y = F_z = 0$, and we have a useful integral relationship from Equation (4):

$$\int_A (B_x^2 + B_y^2) dx dy - \int_A B_z^2 dx dy = 0. \quad (5)$$

Returning to Equation (1), consider the integral

$$\int_V \mathbf{r} \cdot \mathbf{J} \times \mathbf{B} dV = \frac{1}{\mu_0} \int_V \mathbf{r} \cdot (\nabla \times \mathbf{B}) \times \mathbf{B} dV.$$

Transforming the right-hand side, this equation can be written

$$\int_V \mathbf{r} \cdot \mathbf{J} \times \mathbf{B} dV = \frac{1}{\mu_0} \int_V \left[\frac{1}{2} B^2 + \nabla \cdot (\mathbf{B} \mathbf{B} \cdot \mathbf{r} - \frac{1}{2} B^2 \mathbf{r}) \right] dV.$$

If the field is force free, then

$$\frac{1}{2\mu_0} \int_V B^2 dV + \frac{1}{\mu_0} \int_S (\mathbf{B} \mathbf{B} \cdot \mathbf{r} - \frac{1}{2} B^2 \mathbf{r}) \cdot d\mathbf{S} = 0.$$

But the first term is just the magnetic energy E_M , so that

$$E_M = \frac{1}{2\mu_0} \int_S B^2 \mathbf{r} \cdot d\mathbf{S} - \frac{1}{\mu_0} \int_S \mathbf{B} \cdot \mathbf{r} \mathbf{B} \cdot d\mathbf{S}.$$

Again, if we assume the field goes to zero over the surface S except on the photospheric area A , we have

$$E_M = \frac{1}{\mu_0} \int_A B_z (x B_x + y B_y) dx dy.$$

Thus, if the field is force free, the total magnetic energy E_M can be calculated solely from the components of the vector magnetic field in the photosphere.

The energy equation can also be used to test the validity of the force free assumption. When the surface integral is evaluated with different origins for the x, y coordinate system, the value obtained for the energy E_M must be the same if the field is truly force free. If the discrepancies are much greater than any variation expected because of noise in the observational data, one can assume that the overall field of the active region deviates from a force-free configuration.

Using the integral equation, Equation (5), one can assess how well the photospheric vector field was measured. We know

that the line-of-sight field is measured more accurately - we have suggested that the SAMEX magnetograph can measure line-of-sight fields to an accuracy of 1 G whereas we expect a sensitivity of 30 to 50 G for the transverse component. Thus, transverse fields below the 30-50 G level of sensitivity will be lost in the noise. To estimate how much of the actual transverse field is lost, we can use Equation (5) which yields a general relation for non-linear, force-free fields having net zero flux over the region of interest in the x,y plane:

$$\int_A \int B_L^2 dx dy = \int_A \int B_T^2 dx dy,$$

i.e., the integrated square of the line-of-sight field is equal to the integrated square of the transverse field on the photospheric surface. While we may not be resolving all of the line-of-sight field with 0.5" spatial resolution, the measurement of B_L is better than that of B_T , as indicated above. Thus, if this equality is shown to hold (within the errors of the observational data), we have some confidence that we are observing the bulk of the magnetic field.

For fields that are not force free, Equations (2)-(4) indicate that measurements of the photospheric vector field can give us a lower limit on the magnetic forces acting within the volume V above the photosphere. Using these three equations, the magnetic torques \vec{Q} can also be calculated from the relation $\vec{Q} = \vec{r} \times \vec{F}$. These equations will allow us to study the relationship between photospheric forces and torques and the motions and evolution of magnetic features in the photosphere.

To summarize this section, there are several relevant problems that can be addressed at least in part by analyses using these magnetostatic models:

- (1) To what degree is the magnetic field non-potential on a global scale as determined by the data?
- (2) Where are the most non-potential features located within an active region? What areas do they cover? What evolutionary process generated these features?
- (3) What is the three-dimensional character of the field and electric currents in the non-potential features?

These questions and others like them can be addressed with analytical techniques that use potential and linear force-free fields computed from the observed photospheric magnetic field in active regions.

In Table 78 the different extrapolation methods currently in use at MSFC are listed. The computer codes for these methods generally calculate the field components, field lines, and

magnetic energy, using the line-of-sight magnetic field as the lower boundary condition. When the transverse field is calculated using a potential field model and compared with the observed transverse field component, the comparison will indicate the deviation of the real field from the assumed potential field configuration, i.e., the non-potential characteristics of an active region. In a similar manner, a comparison of computed and observed transverse fields can be made assuming a force-free field with a chosen value of the parameter alpha; the extent of agreement will indicate how well the constant value of alpha models the real field throughout the active region. The linear force-free methods can also be used to calculate a potential field; calculations of the potential field by different methods can be used to check the constraints of the boundary conditions of each method on the extrapolated field.

Table 78. MSFC programs to extrapolate photospheric magnetic fields into the transition region and corona.

Method	Calculation	Computed Quantities		
		Components	Field Lines	Energy
Schmidt-Harvey	Potential	x	x	x
Teuber Potential	x		x	
Nakagawa-Raadu-Welck	Linear force free	x	x	x
Alissandrakis	Linear force free	x		x
Wu-Chang-Hagyard	Non-linear, force free	x	x	x

3.2 Parametric Studies

Generally speaking, observational techniques in solar astronomy are far ahead of our ability to provide a theoretical description of the phenomena observed. In instances where this is true, parametric studies of observed properties can provide insight into what physical processes are important in the theory and thus what assumptions are valid. In this respect, observed parameters can influence the development of theoretical models. These parametric classifications also produce a form of synopsis of the observational data that is useful for the theorist, providing him further insight into the observational data. Phenomenological parameters derived from the

SAMEX magnetograph include: measured flux, field strengths, nature of the neutral line, magnetic shear, areal extent of shear, growth rates, decay rates, magnetic gradients, potential and force-free characteristics of the field, estimates of magnetic energy and Lorentz forces and torques, and electric currents. Statistical analyses of these parameters need to be developed for flaring and non-flaring active regions and for active regions at different times in the solar cycle. The use of parametric studies will also be important in developing techniques to predict the location and onset of flares.

3.3 Magnetohydrodynamical Models

The general equations of magnetohydrodynamics (MHD) can be written (Priest, 1982):

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{v} \times \mathbf{B}) + \eta \nabla^2 \mathbf{B},$$

$$\frac{D\rho}{Dt} + \rho \nabla \cdot \mathbf{v} = 0,$$

$$\rho \frac{D\mathbf{v}}{Dt} = -\nabla p + \mathbf{J} \times \mathbf{B} + \mathbf{F},$$

$$p = \frac{k}{m} \rho T.$$

In these equations, \mathbf{v} is the plasma velocity, ρ is mass density, k is the Boltzmann constant, p is the plasma pressure, η is the magnetic diffusivity, and \mathbf{F} represents the sum of viscous and gravitational forces. The material derivative for time variations following the motion is defined by

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + \mathbf{v} \cdot \nabla.$$

These equations couple the velocity \mathbf{v} with the magnetic field \mathbf{B} . The current density \mathbf{J} is determined by the magnetic field via the relation

$$\mu_0 \mathbf{J} = \nabla \times \mathbf{B}.$$

In addition, the magnetic induction has zero divergence:

$$\nabla \cdot \mathbf{B} = 0.$$

Mathematically, this system of MHD equations forms an initial boundary value problem, in which the initial boundary conditions are obtained from observations. Specifically, the initial magnetic field configuration used as input to the MHD model simulating an active region could be obtained from observations with the SAMEX

magnetograph. This is in contrast to earlier studies by Wu (1983a, 1986b) or the numerical simulations of Van Hoven et al. (1981) showing the response of the magnetic field to photospheric driving velocities in which hypothetical magnetic field configurations were always used as the initial conditions. With the high spatial resolution measurements of both the velocity and magnetic fields that will be obtained with the SAMEX magnetograph, effects of local mass motion coupled with the magnetized plasma can be investigated realistically for the first time.

Because of this coupling between the magnetic field, currents, and velocity, the motion of the solar plasma in the magnetic field is an important factor in the evolution of active regions. Photospheric shearing motions and proper motions of sunspots generate interactions among magnetic flux tubes which can lead to field reconnections and the formation of current sheets. With the SAMEX observations, the production of current sheets may be observed for the first time using cotemporal measurements of the magnetic field and x-ray images.

Time-dependent magnetohydrodynamical models would be used to study this evolution of solar active regions and the process of flare energy buildup and release. Current models include simple one-dimensional, time-dependent hydrodynamic models (Nakagawa et al., 1975; Wu et al., 1975, 1976) and multi-dimensional models (Wu et al., 1983a,b,c, 1984, 1985b, 1986a,b, 1987a; Dryer et al., 1986).

The MHD simulations developed by Wu and his collaborators model physical processes from the solar surface to the Earth's upper atmosphere, starting with an observed shearing motion at the photosphere and simulating the spatial and temporal evolution of the ensuing disturbance all the way to the Earth. These models, illustrated in Figure 74, are summarized as follows:

(1) Two-dimensional, time-dependent, non-planar, ideal MHD model in rectangular coordinates; the model extends from the solar photosphere to an altitude of 30,000 km and is used to model the energy buildup prior to a solar flare.

(2) Two-dimensional, time-dependent, non-planar, ideal MHD model in spherical axisymmetric coordinates; the model uses an isothermal hydrostatic atmosphere from 20,000 km altitude to a distance of 10 solar radii and is used to study coronal dynamics.

(3) Two-dimensional, time-dependent, non-planar, ideal MHD model in spherical axisymmetric coordinates with an MHD equilibrium atmosphere; the model starts with a uniform, steady-state solar wind as its initial state and is used to study interplanetary dynamics.

These ideal MHD models employ the simplest energy equations and thus ignore some important physical mechanisms. Recent advances in numerical techniques (Hu and Wu, 1984) and the advent of increased computing power will enable Wu and his collaborators to include more realistic physics into their models:

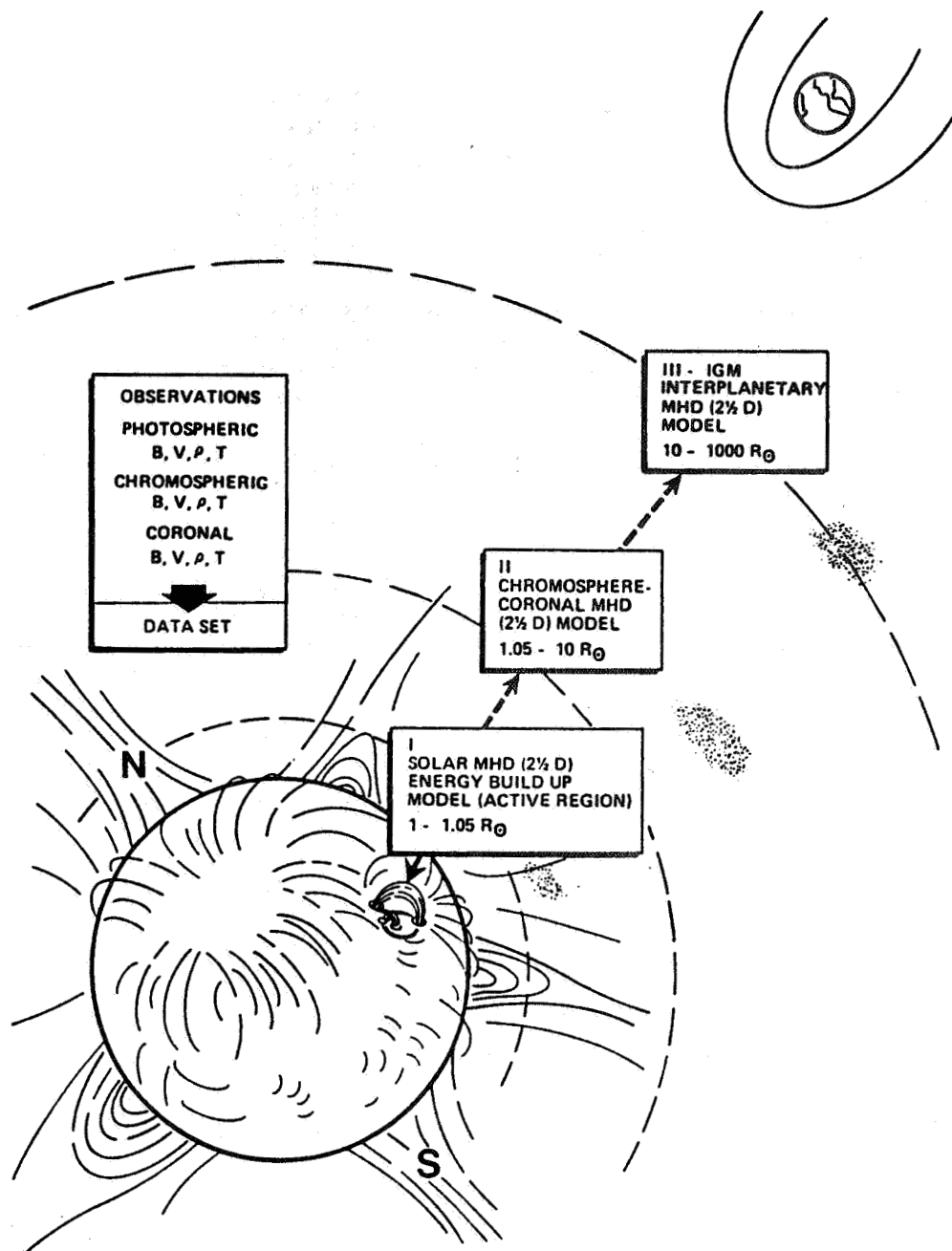


Figure 74. Schematic representation of the regime of solar-interplanetary models. MHD models developed by S. T. Wu and his collaborators simulate physical processes from the solar surface to the Earth's upper atmosphere. This figure shows the three regimes of the Sun-Earth space that are modeled.

(1) MHD models with realistic energy equation - two-dimensional, time-dependent, non-planar model with finite electrical conductivity and thermal conduction; the models will study flare energy buildup and release caused by shearing motions, compressible magnetic reconnection processes in relation to flare triggering, and coronal heating due to shearing or twisting motions (Parker, 1981).

(2) Three-dimensional magnetohydrostatic equilibrium models for prominences, the solar corona, and coronal loops; perturbed solutions will be sought to study initiation of coronal mass ejections and the response to them by the heliosphere.

(3) Extension of the three-dimensional MHD model using explicit numerical techniques for supersonic and super-Alfvénic flow to study traveling interplanetary disturbances.

These computational models will all play major roles in the data analysis and modeling programs associated with the SAMEX mission. Selected data from the SAMEX instruments and from other solar observations will be combined with these numerical models to carry out numerical experiments on solar activity. Observed configurations of the photospheric magnetic field and inferred plasma properties will be used as inputs to the inner coronal model (Wu et al., 1983a) to compute buildup and release of flare energy. The computed energy released will then be used as input for the coronal MHD model (Wu et al., 1983b) which extends the calculation to about 10 solar radii. At this point, the interplanetary model (Wu et al., 1983c) will be used to propagate the disturbances to the Earth's environment.

Using the SAMEX data in such a scenario of MHD computations, we will be able to model many of the processes which give rise to solar activity and some of the terrestrial effects produced by this activity, with the goal of trying to advance our understanding of the physical mechanisms involved in this complex coupling between Sun and Earth. In so doing, one of the primary goals of the SAMEX mission will be achieved.

VI. CONCLUSIONS AND RECOMMENDATIONS

1. Science Summary

The SAMEX instruments have been selected and designed to specifically study the role of the Sun's magnetic field in solar activity, starting with the emergence of the magnetic field in the photosphere and following its evolution through the most dynamic stages of its development until it finally dissipates and disappears. In Figure 75 (after Neidig, 1986) the processes that define this activity and that will be studied with the SAMEX instruments are schematically summarized.

As with all scientific investigations that push forward the frontier of a particular discipline, old problems may be solved or intensified while totally new ones will inevitably be uncovered. Dr. D. M. Rabin of the National Solar Observatory has summarized this aspect of the SAMEX mission in his contribution to the SAMEX Science Meeting sponsored by the High Altitude Observatory in September 1986. We present his summary in sections VI.1.1-VI.1.3 as an overview of our expectations for the scientific dividends of SAMEX.

1.1 Problems SAMEX Will Solve

We should expect to solve a number of important and longstanding problems in solar physics, where by "solve" we mean asking new questions rather than repeating old ones. These solar problems include:

- (1) Statistical relationships with predictive value between the occurrence of solar flares and: magnetic shear, electric currents and bulk flows in the photosphere; magnetic energy buildup; emerging and submerging flux; coronal precursors.
- (2) Magnetic field configuration and physical driver of spicules.
- (3) Properties of electric currents crossing the photosphere in active regions and flowing in the atmosphere above the photosphere.
- (4) Gross structure of sunspot magnetic fields.
- (5) Structure of the magnetic field around and over the network.
- (6) "Direct" evidence for emerging and submerging magnetic fields.
- (7) Relative importance of magnetic diffusion and in situ removal of photospheric flux in active regions.

ACTIVE REGION FLARE AND PRE-FLARE PHENOMENA

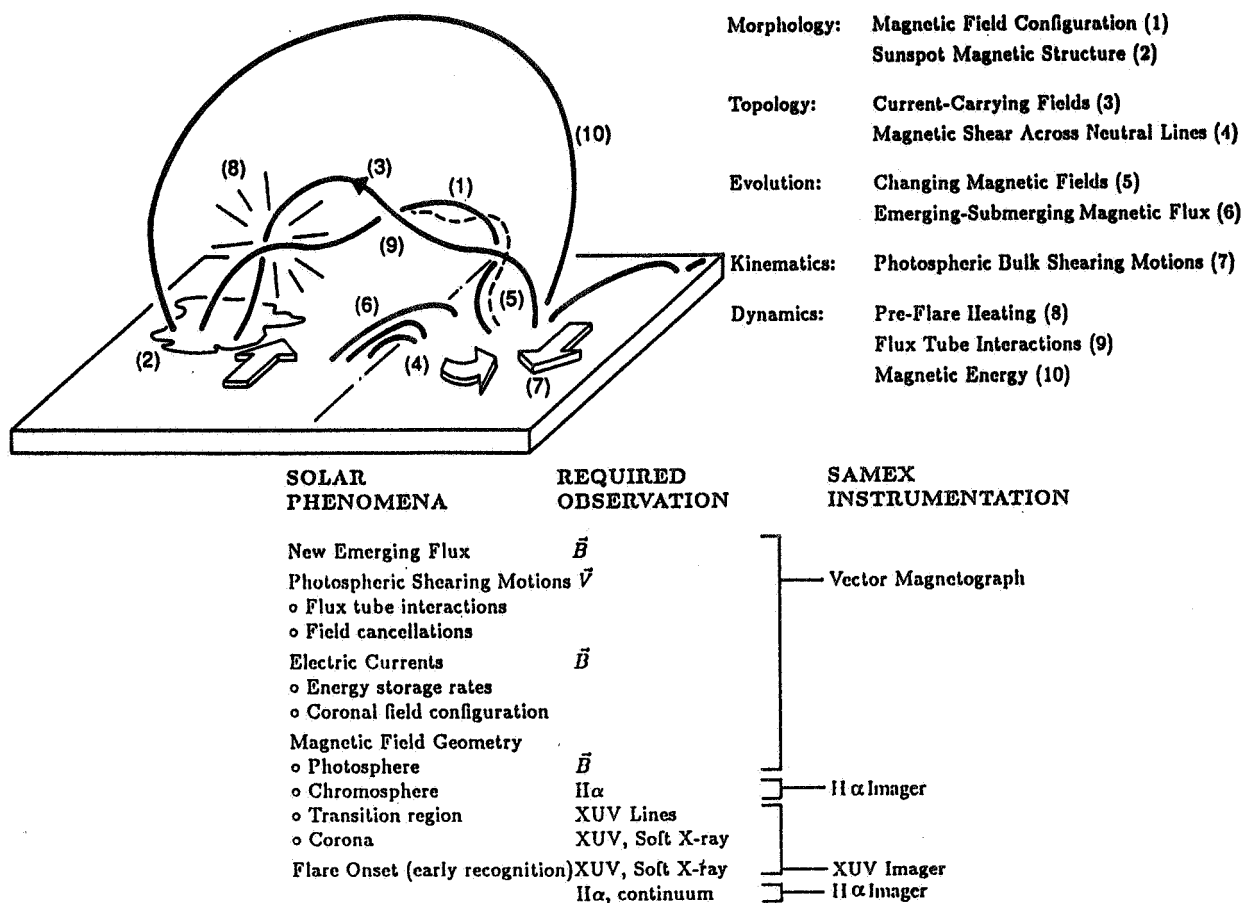


Figure 75. Schematic representation of physical processes studied with the SAMEX instruments. The specific solar phenomena associated with flares and the pre-flare state that will be studied with the SAMEX instruments are listed with the observational requirements and the relevant SAMEX instrument. In the representation of an active region, magnetic flux tubes are shown emerging, moving, changing, and interacting to produce the pre-flare state and the flare event. The major role and importance of the solar magnetic field will be studied in the photosphere, chromosphere, transition region, and corona. Such a study of the morphology, topology, evolution, kinematics, and dynamics of active regions will reveal the detailed processes which give rise to solar activity. SAMEX will make coordinated high-resolution observations of the magnetic and dynamical characteristics of solar active-region development, flares, and pre-flare energy buildup using the vector magnetograph described in this report. Simultaneous observations with the XUV imager will provide the relationship between magnetic fields and the high temperature coronal plasma structures and instabilities.

- (8) Properties of network and internetwork magnetic loops.

1.2 Problems SAMEX Will Partially Solve

SAMEX will also help to find some of the answers to solar problems such as:

- (1) The origin of magnetic shear and the loss of equilibrium in non-potential (sheared) configurations.
- (2) The existence of fine-scale electric currents and joule heating in the upper atmosphere.
- (3) Magnetic "burning" versus eruptive microflares - is wave heating required in coronal holes?
- (4) Lower-atmospheric signatures of coronal mass ejections.
- (5) High temperature structure in flares.
- (6) The origin of the fine-scale structure in the upper atmosphere: how much can be mapped downward to photospheric structure? How much is produced in situ, and how (reconnection? thermal instability? electrothermal instability? wave propagation?)? Why are active regions more inhomogeneously filled than quiet regions?
- (7) How do large- and small-scale features interact?

1.3 Problems SAMEX Will Intensify

- (1) All the questions SAMEX will partially solve!
- (2) Solar cycle variations.
- (3) Spatial structure not fully resolved.
- (4) Temporal structure not fully resolved.
- (5) The relationship between photospheric, transitional, and coronal velocity fields.
- (6) The relationship between coronal and transitional fine structures.
- (7) Physical conditions within coronal fine structures.
- (8) Physical conditions within magnetic flux tubes.

- (9) Interpretation of vector magnetograph data!
- (10) The flare mechanism.
- (11) Balancing the books on flare energetics.

2. Instrument Summary

It is no accident of arrangement that this SAMEX report begins with a science discussion and a presentation of the scientific objectives of the SAMEX mission. The science was (and should continue to be) the driving force in defining the specific design of the SAMEX instruments. Using the science objectives defined for the mission, we developed five primary instrument requirements for the SAMEX vector magnetograph that are driven by the SAMEX science:

- o the highest possible sensitivity in measuring the transverse component of the vector magnetic field in the photosphere;
- o a spatial resolution of 0.5 arc sec;
- o a temporal resolution of a few minutes;
- o a field of view large enough to map most active regions;
- o a spectral range sufficient to permit flux tube diagnostics.

Of these requirements, the primary driver was magnetic sensitivity since so many of the important physical parameters involved in solar activity are derived from the magnetic field measurements and need to be specified as accurately as possible.

To meet these requirements within the constraints of a space mission, we were forced to conduct various trade-off studies to arrive at the final design that is presented in this report. A significant factor in many decisions was the desire to keep the design as simple as possible, although we knew that a vector magnetograph is an extremely complex instrument even as a ground-based system. However, throughout this design effort, the bottom line was magnetic sensitivity so that the primary goal in the design was the measurement of polarized intensities with the highest accuracy that modern technology permits. Consequently, the development of the polarimeter design and the analysis of cross talk and instrumental polarization were done in great detail to achieve this goal.

Overall, the design is quite feasible. However, as with any design, the results of theoretical analyses need to be verified experimentally, and there are certain areas where technological development is either required or desirable. To

address these areas of concern, in the following sections that end this report we give our recommendations for follow-on efforts to this design study.

3. Recommended Studies

There are areas in the SAMEX science, instrumentation, and mission operations that need further development before a final mission concept is derived. While the list we present here is not intended to be complete, it does highlight major areas uncovered in the course of this study.

3.1 Theoretical Foundations

Further development of the theories relating magnetic fields to solar activity needs to occur prior to the SAMEX mission. Specific examples include instabilities leading to flare onset, magnetic reconnection processes, and non-linear force-free field calculations.

3.2 Instrument Design

While the optical design presented in this report is very detailed and was developed through extensive effort, we feel that it can be "fine tuned" to produce an even more optimum design. This fine-tuning would be done to simplify the optics, shorten the optical train, and reduce even further the instrumental polarization. With respect to this last goal, more research needs to be done on the designs of the optical coatings for the telescope mirrors and the optical lenses. The basic computer programs for the optical and coating designs were developed for this study; they would be used to analyze the effects of different materials on the various optical components and how each coating interfaces with the other components in the optical train to minimize the polarization effects while optimizing the optical transmission.

3.3 Mission Operations

A decision will have to be made very early in the planning stage about the orbit of the SAMEX spacecraft since this affects so many aspects of the mission. We have presented the various options in this report and indicated that the selection of orbit should be driven by the observational requirements imposed by the scientific objectives.

Because of the severe limitations imposed by the restricted telemetry from most orbits, a large effort is needed for planning the various coordinated observing modes of all three of the SAMEX instruments. This advance planning is necessary to optimize the amount of science returned with the limited amount of telemetry available.

4. Recommended Test Programs

Experimental tests should be carried out to confirm analytical results developed in the design of the optics, polarimeter, and spectral filters, and to verify that the instrumental polarization of the design can be reduced to acceptable levels.

4.1 Optical Testing

The following tests of the optical design are recommended:

- (1) Test all optics for ghost images and their effects.
- (2) Evaluate effects of off-set pointing on optical performance.
- (3) Evaluate effects of scattered light on optical performance.
- (4) Test tolerance of optics for space flight conditions.

4.2 Polarimeter Testing

The following tests are recommended to verify the design of the polarimeter proposed in this document:

- (1) Construct a breadboard system.
- (2) Test performance of breadboard system.
- (3) Test interface of breadboard with other optical components.

4.3 Tests of Spectral Filters

Both a birefringent and Fabry-Perot filter should be tested in the designed optical system. Two problems need to be addressed specifically: (1) the location of the filter in the optical path (telecentric versus collimated system) and how this location affects the polarization accuracy; (2) how the filter interfaces with the polarimeter.

4.4 Tests of Instrumental Polarization

Our analysis of instrumental polarization for the optical design indicated that we can achieve a reduction of the instrumental polarization to about 10^{-5} , the design goal for the optical system. These results need to be experimentally verified. Some recommended tests include:

- (1) Coat mirrors and lenses with computer-designed coatings and test them for instrumental polarization.

- (2) Analyze effects of scattered light on polarization accuracy.
- (3) Analyze effects of off-set pointing of the secondary mirror on polarization accuracy.

5. Recommended Developments in Technology

We feel that there will have to be some advanced development in certain areas of technology to realize this design for the SAMEX magnetograph. However, none of this development requires any significant breakthrough. The requirement generally is the application of present technology in new areas or in different ways. Specific areas include:

- (1) Development of space-qualified computers (MicroVAX class), array processors, mass storage units, and microprocessors.
- (2) Development of improved telemetry techniques.
- (3) Development of large-array, solid-state detectors with rapid readout architecture and high signal-to-noise ratio.

6. Conclusion

The study we have completed indicates that it is entirely feasible to build a vector magnetograph that fulfills the requirements driven by the scientific objectives of the SAMEX mission. To achieve the goal, however, will require further development and testing as outlined in the previous sections. The time is at hand to start this preliminary work and proceed to the development of the flight instrument. Only with these careful and diligent efforts now will we attain the ultimate goal of orbiting the most accurate vector magnetograph that modern technology can produce to discover and explore a revolutionary new magnetic Sun.

REFERENCES

- Aikens, R. S., Lynds, C. R., and Nelson, R. E.: 1976, Proc. SPIE **78**, 65.
- Akiyama, I., Tanaka, T., Oda, E., Kamata, T., Masubuchi, K., Arai, K., and Ishihara, Y.: 1986, IEEE International Solid-State Circuits Conference, p. 96.
- Allen, C. W.: 1973, Astronomical Quantities, 3rd Edition, Athlone Press, London, p. 113.
- Allen, L. B.: 1972, DTIC Tech. Rep. AFAL-TR-72-60.
- Alissandrakis, C. E.: 1981, Astron. Astrophys. **100**, 197.
- Aly, J. J.: 1984, Astrophys. J. **283**, 349.
- Athay, R. G., Jones, H. P., and Zirin, H.: 1985, Astrophys. J. **291**, 344.
- Athay, R. G., Klimchuk, J. A., Jones, H. P., and Zirin, H.: 1986, Astrophys. J. **303**, 884.
- Barbe, D. F.: 1975, Proc. IEEE **63**, 38.
- Barbe, D. F., and Campana, S. B.: 1976, in P. G. Jespers, F. Van de Wiele, and M. H. White (eds.), Solid State Imaging, Noordhoff, Leyden, p. 623.
- Bar-lev, A.: 1979, Semiconductors and Electronic Devices, Prentice/Hall International, p. 32.
- Beckers, J. M.: 1969, Solar Phys. **9**, 372.
- Beckers, J. M.: 1971, in R. Howard (ed.), Solar Magnetic Fields, IAU Symp. 43, Reidel, Dordrecht, p. 3.
- Bennett, J. M., and Bennett, H. E.: 1978, in Walter G. Driscoll (ed.), Handbook of Optics, McGraw-Hill, New York, pp. 10-30.
- Berry, R. W., Hall, P. M., and Harris, M. T.: 1968, Thin Film Technology, D. Van Nostrand Co., Inc., p. 330.
- Berton, R., and Sakurai, T.: 1985, Solar Phys. **96**, 93.
- Birn, J., Goldstein, H., and Schindler, K.: 1978, Solar Phys. **57**, 81.
- Blouke, M. M., Janesick, J. R., Hall, J. E., and Cowens, M. W.: 1981, Proc. SPIE **290**, 6.

- Blouke, M. M., Heidtmann, D. L., Corrie, B., Lust, M. L., and Janesick, J. R.: 1985, Proc. SPIE **570**, 82.
- Borsky, D.: 1981, Electronic Design **29**, 81.
- Bower, F. H.: 1984, Fairchild Charge Coupled Device Catalog, p. 173.
- Brants, J. J.: 1985, Solar Phys. **98**, 197.
- Brown, D. M., Burke, H. K., Ghezze, M., McConnolee, P., Michon, G., and Vogelsong, T. L.: 1980, IEEE International Solid-State Circuits Conference, p. 28.
- Burke, H. K., and Michon, G. J.: 1976, IEEE Trans. Electron Devices **23**, 189.
- Carnes, J. E., and Kosonocky, W. F.: 1972, RCA Review **33**, 327.
- Chipman, R.: 1987, "Polarization Aberrations," Ph.D. Thesis, University of Arizona, Tucson, Arizona (in press).
- Chiu, Y. T., and Hilton, H. H.: 1977, Astrophys. J. **212**, 873.
- De Jager, C.: 1986, Adv. Space Res. **6(6)**, 353.
- deLoach, A. C., Hagyard, M. J., Rabin, D., Moore, R. L., Smith, J. B., West, E. A., and Tandberg-Hanssen, E.: 1984, Solar Phys. **91**, 235.
- Dereniak, E. L., and Crowe, D. G.: 1984, Optical Radiation Detectors, John Wiley and Sons, New York, p. 186.
- Deyhimy, I., Harris, J. S., Jr., Eden, R. C., Anderson, R. J., and Edwall, D. D.: 1979, IEEE Trans., p.26.4.
- Dowdy, J. F., Jr., Rabin, D., and Moore, R. L.: 1986, Solar Phys. **105**, 35.
- Dryer, M., Smith, Z. K., Wu, S. T., Han, S. M., and Yeh, T.: 1986, in Y. Kamide and J. A. Slavin (eds.), Solar Wind-Magnetosphere Coupling, AGU Chapman Conference, Terra Scientific Publ. Co., Tokyo.
- Eddy, J. A.: 1979, A New Sun (The Solar Results from Skylab), NASA SP-40, p. 122.
- Ellison, M. A.: 1955, The Sun and Its Influence, MacMillan, New York, p. 117.
- Gaizauskas, V., and Harvey, K. L.: 1986, Adv. Space Res. **6(6)**, 17.
- Gaizauskas, V.: 1986, Solar Phys. **105**, 67.
- Gary, G. A., Moore, R. L., Hagyard, M. J., and Haisch, B. M.: 1987, Astrophys. J. **314**, 782.

- Gesztelyi, L., Karlicky, M., Farnik, F., Gerlei, O., and Valnicek, B.: 1986, in D. F. Neidig (ed.), The Lower Atmosphere of Solar Flares, p. 163.
- Gibson, E. G.: 1973, The Quiet Sun, NASA SP-303, p. 127.
- Giovanelli, R. G., and Jones, H. P.: 1982, Solar Phys. **79**, 267.
- Glass, A. M., Peterson, G. E., and Negran, T. J.: 1972, in Laser Induced Damage in Optical Materials, NBS Publ. 372, p. 15.
- Green, J. L., and Peters, D. J.: 1985, Introduction to the Space Plasma Physics Analysis Network (SPAN) - First Edition, NASA Tech. Memorandum TM-86499.
- Gunn, J. E., and Westphal, J. A.: 1981, Proc. SPIE **290**, 16.
- Gunning, W., Pasko, J., and Tracy, J.: 1981, Proc. SPIE **268**, 190.
- Gunning, W.: 1982, Appl. Opt. **21**, 3129.
- Hagyard, M., Low, B. C., and Tandberg-Hanssen, E.: 1981, Solar Phys. **73**, 257.
- Hagyard, M. J., and West, E.: 1982, MSFC Internal Note.
- Hagyard, M. J., Cumings, N. P., West, E. A., and Smith, J. E.: 1982, Solar Phys. **80**, 33.
- Hagyard, M. J., Moore, R. L., and Emslie, A. G.: 1984a, Adv. Space Res. **4**, 71.
- Hagyard, M. J., Smith, J. B., Jr., Teuber, D., and West, E. A.: 1984b, Solar Phys. **91**, 115.
- Hagyard, M. J., Cumings, N. P., and West, E. A.: 1985a, in C. De Jager and Chen Biao (eds.), Proceedings of Kunming Workshop on Solar Physics and Interplanetary Travelling Phenomena, p. 1216.
- Hagyard, M., West, E. A., and Smith, J. B., Jr.: 1985b, in C. De Jager and Chen Biao (eds.), Proceedings of Kunming Workshop on Solar Physics and Interplanetary Travelling Phenomena, p. 179.
- Hagyard, M. J., and Rabin, D. M.: 1986, Adv. Space Res. **6(6)**, 7.
- Haisch, B. M., Bruner, M. E., Hagyard, M. J., and Bonnet, R. M.: 1986, Astrophys. J. **300**, 428.
- Hannakam, L., Gary, G. A., and Teuber, D. L.: 1984, Solar Phys. **94**, 219.
- Harris, S. E., and Wallace, R. W.: 1969, J. Opt. Soc. Amer. **59**, 744.

- Harvey, J. W.: 1966, HAO Astro-Geophysical Memo. 172.
- Harvey, J. W.: 1969, NCAR Cooperative Thesis No. 17, University of Colorado and High Altitude Observatory, NCAR.
- Harvey, J.: 1977, in E. A. Muller (ed.), Highlights of Astronomy 4, Part II, p. 223.
- Harvey, J.: 1987, in GONG Newsletter No. 3.
- Harvey, K. L., Livingston, W. C., Harvey, J. W., and Slaughter, C. D.: 1971, in R. Howard (ed.), Solar Magnetic Fields, IAU Symp. 43, Reidel, Dordrecht, p. 422.
- Heyvaerts, J., Priest, E. R., and Rust, D. M.: 1977, Astrophys. J. 216, 123.
- Hogan, B. J.: 1986, Design News, March 17 Issue, p. 72.
- Holman, G. D., and Kundu, M. R.: 1985, Astrophys. J. 292, 291.
- Howard, R., and LaBonte, B. J.: 1981, Solar Phys. 74, 131.
- Hu, Y. Q., and Wu, S. T.: 1984, J. Comp. Phys. 55, 33.
- Hynecek, J.: 1979, Virtual Phase CCD Technology, IEEE Conf., p. 611.
- Ishirara, Y., Takeuchi, E., Teranishi, N., Kohono, A., Aizawa, T., Arai, K. and Shiraki, H.: 1980, IEEE International Solid-State Circuits Conference, p. 24.
- JHU/APL: 1985, The Johns Hopkins University Applied Physics Laboratory Doc. SDO 7805, p. 156.
- Jorden, P. R., and van Breda, I. G.: 1981, Proc. SPIE 290, 113.
- Kjeldseth-Moe, O.: 1968, Solar Phys. 4, 267.
- Kosonocky, W. F., and Carnes, J. E.: 1973, RCA Review 34, 164.
- Krall, K. R., Smith, J. B., Jr., Hagyard, M. J., West, E. A., and Cumings, N. P.: 1982, Solar Phys. 79, 59.
- Krieger, A. S., Vaiana, G. S., and Van Speybroeck, L. P.: 1971, in R. Howard (ed.), Solar Magnetic Fields, IAU Symp. 43, Reidel, Dordrecht, p. 397.
- Kubota, H., and Inoue, S.: 1959, J. Opt. Soc. Am. 49, 191.
- Landi Degl'Innocenti, E.: 1983, Solar Phys. 85, 3.
- Levine, R. H.: 1975, Solar Phys. 44, 365.

- Levine, R. H.: 1976, Solar Phys. **46**, 159.
- Light, W.: 1982, Non-Contact Position Sensing Using Silicon Detectors, United Detector Technology, Hawthorne, California.
- Lin, R. P., Schwartz, R. A., Pelling, P. M., and Harley, K. C.: 1981, Astrophys. J. (Lett.) **151**, L109.
- Lingren, R. W., and Tarbell, T.D.: 1981, Proc. SPIE (Suppl.) **265**, 47.
- Livingston, W. C.: 1974, in Y. Nakagawa and D. M. Rust (eds.), Flare Related Magnetic Field Dynamics, Natl. Ctr. Atmos. Res., Boulder, Colorado, p. 269.
- Low, B. C.: 1977a, Astrophys. J. **212**, 234.
- Low, B. C.: 1977b, Astrophys. J. **217**, 988.
- Low, B. C.: 1982, Solar Phys. **77**, 43.
- Low, B. C.: 1985, in M. J. Hagyard (ed.), Measurements of Solar Vector Magnetic Fields, NASA CP-2374, p. 49.
- Luppino, N., Ceglio, M., Doty, J. P., Ricker, G. R., and Vallergera, J. V.: 1986, Proc. SPIE X-ray Imaging II.
- Lyot, B.: 1944, Ann. Astrophys. **7**, 31.
- Machado, M. E., and Moore, R. L.: 1986, Adv. Space Res. **6(6)**, 217.
- Martin, P. J.: 1986, J. Mat. Sci. **21**, 1.
- Martin, S. F., Livi, S. H. B., and Wang, J.: 1985, Aust. J. Phys. **38**, 929.
- Martin, S. F.: 1986, in A. I. Poland (ed.), Coronal and Prominence Plasmas, NASA CP-2442, p. 73.
- McGrath, R. D., Freeman, J. W., and Janesick, J.: 1983, IEEE Trans., p. 19.1.
- Molodensky, M. M.: 1974, Solar Phys. **39**, 393.
- Moore, R. L., Tang, F., Bohlin, J. D., and Golub, L.: 1977, Astrophys. J. **218**, 286.
- Moore, R. L., and LaBonte, B. J.: 1980, in M. Dryer and E. Tandberg-Hanssen (eds.), Solar and Interplanetary Dynamics, IAU Symp. 91, Reidel, Dordrecht, p. 207.
- Moore, R. L.: 1981, in L. E. Cram and J. H. Thomas (eds.), The Physics of Sunspots, Sacramento Peak Observatory Conference, p. 259.

- Moore, R. L., Hurford, G. J., Jones, H. P., and Kane, S. R.: 1984, Astrophys. J. **276**, 379.
- Moore, R., and Rabin, D.: 1985, Ann. Rev. Astron. Astrophys. **23**, 239.
- Moore, R. L., Kahler, S. W., Kane, S. R., and Zirin, H.: 1986, Bull. AAS **18**, 708.
- Moreton, G. E., and Severny, A. B.: 1968, Solar Phys. **3**, 282.
- Mount Stromlo and Siding Spring Observatory: 1981, Preliminary Design and Development Report for Starlab Instrument Package, Woden, Australia.
- Nakagawa, Y., and Raadu, M. A.: 1972, Solar Phys. **127**, 135.
- Nakagawa, Y., Wu, S. T., and Tandberg-Hanssen, E.: 1975, Solar Phys. **41**, 387.
- Neidig, D. F. (ed.): 1986, Scientific Uses of the SAMEX Satellite System, Report on the SAMEX Science Meeting, September 1986, Boulder, Colorado.
- Neidig, D. F., Hagyard, M. J., Smith, J. B., Jr., and Machado, M. E.: 1986, Adv. Space Res. **6**(6), 25.
- Orrall, F. Q. (ed.): 1981, Solar Active Regions, Colorado Associated University Press, Boulder, Colorado.
- Parker, E. N.: 1975, Astrophys. J. **198**, 205.
- Parker, E. N.: 1979, Cosmical Magnetic Fields, Clarendon Press, Oxford, p. 739.
- Parker, E. N.: 1981, Astrophys. J. **244**, 631.
- Porter, J. G., Toomre, J., and Gebbie, K. B.: 1984, Astrophys. J. **283**, 879.
- Porter, J. G., Reichmann, E. J., Moore, R. L., and Harvey, K. L.: 1985, Bull. AAS **17**, 842.
- Porter, J., Reichmann, E., Moore, R., and Harvey, K.: 1986, in A. I. Poland (ed.), Coronal and Prominence Plasmas, NASA CP-2442, p. 383.
- Porter, J. G., Moore, R. L., Reichmann, E. J., Engvold, O., and Harvey, K. L.: 1987, Astrophys. J. **323**, 380.
- Pridmore-Brown, D. C.: 1981, Rpt. ATR-81(7813)-1, Aerospace Corporation, El Segundo, California.

- Priest, E. R.: 1982, Solar Magnetohydrodynamics, Reidel, Dordrecht.
- Priest, E. R.: 1985, Rep. Prog. Phys. **48**, 955.
- Rabin, D. M., R. L. Moore, and M. J. Hagyard: 1984, Astrophys. J. **287**, 404.
- Rabin, D. M.: 1986, in D. F. Neidig (ed.), Scientific Uses of the SAMEX Satellite System, Report on the SAMEX Science Meeting, September 1986, Boulder, Colorado.
- Reddersen, B. R., and Ralston, L. M.: 1980, Optical Eng. **19**, 199.
- Robinson, A. L.: 1987, Science **235**, 1324.
- Rosenberg, W. J., and Title, Alan M.: 1981, AGARD-CP-300 Spec. Topics in Opt. Propagation, Advisory Group for Aerospace Res. and Dev. Conf. Proceedings No. 300, Neuilly Sur Seine, France, p. 32-1.
- Rosenberg, W. J.: 1986, Compact Photometric and Imaging Spectroscopic System, Final Report, NASW-3662 (LMSC-F108007).
- Rust, D. M.: 1973, Solar Phys. **33**, 205.
- Rust, D. M.: 1985, in M. J. Hagyard (ed.), Measurements of Solar Vector Magnetic Fields, NASA CP-2374, p. 141.
- Rust, D. M., Burton, C. H., and Leistner, A. J.: 1986a, Proc. SPIE **627**, 39.
- Rust, D. M., Appourchaux, T., and Hill, F.: 1986b, in J. Christensen-Dalsgaard (ed.), Advances in Helio- and Astroseismology, IAU Symp. 123.
- Sakurai, T.: 1979, Publ. Astron. Soc. Japan **31**, 309.
- Sakurai, T.: 1981, Solar Phys. **69**, 343.
- Sakurai, T.: 1982, Solar Phys. **76**, 301.
- SAMSAT Pre-Phase A Study: 1984, Program Development Office, MSFC.
- Savoye, E. D.: 1985, Proc. SPIE **570**, 95.
- Sawyer, C., Warwick, J. W., and Dennett, J. T.: 1986, Solar Flare Prediction, Colorado Associated University Press, Boulder, Colorado, p. 93.
- Schadee, A., De Jager, C., and Svestka, Z.: 1983, Solar Phys. **89**, 287.
- Schmidt, H. U.: 1964, in W. N. Hess (ed.), Physics of Solar Flares, NASA SP-50, p. 107.

- Seehafer, N.: 1982, Solar Phys. **81**, 69.
- Sequin, C. H., and Tompsett, M. F.: 1975, Charge Transfer Devices, Academic Press, New York, p. 70.
- Sherrill, T. J.: 1981, AIAA Report AIAA-81-0169, Space Telescope Orbiting Viewing Constraints.
- Shurcliff, W. A.: 1962, Polarized Light, Harvard University Press, Cambridge, Massachusetts, p. 15.
- Simon, P., Smith, J. B., Jr., Ding, Y., Flowers, W., Guo, Q., Harvey, K. L., Hedeman, R., Martin, S. F., McKenna Lawlor, S., Lin, V., Neidig, D., Obridko, V. N., Dodson Prince, H., Rust, D., Speich, D., Starr, A., and Stepanyan, N. N.: 1979, in R. F. Donnelly, (ed.) Solar-Terrestrial Predictions Proceedings, Vol. 2, U.S. Dept. of Commerce, p. 287.
- Smith, J. B., Jr., and Neidig, D. F.: 1986, in P. A. Simon, G. Heckman, and M. A. Shea (eds.), Solar-Terrestrial Predictions, U.S. Dept. of Commerce, P. 167.
- Smith, R. C., and Tarbell, T. D.: 1977, Correlation Tracking Study for a Meter-Class Solar Telescope of the Space Shuttle, Lockheed Palo Alto Research Laboratory, LMSC/D562493.
- Solar-Geophysical Data: 1983, Comprehensive Reports, Number 463, Part II, NOAA, National Geophysical Data Center, Boulder, Colorado.
- Stahl, P. A.: 1983, J. Roy. Astron. Soc. Can. **77**, 203.
- Stenflo, J. O.: 1971, in R. Howard (ed.), Solar Magnetic Fields, IAU Symp. 43, Reidel, Dordrecht, p. 102.
- Stenflo, J. O.: 1973, Solar Phys. **32**, 41.
- Stenflo, J. O. (ed.): 1983, Solar and Stellar Magnetic Fields: Origins and Coronal Effects, Reidel, Dordrecht.
- Stenflo, J. O.: 1985, in M. J. Hagyard (ed.), Measurement of Solar Vector Magnetic Fields, NASA CP-2374, p. 263.
- Stratton, J. A.: 1941, Electromagnetic Theory, McGraw-Hill, New York, p. 97.
- Svestka, Z.: 1976, Solar Flares, Reidel, Dordrecht, p. 310.
- Tanaka, K.: 1978, Solar Phys. **58**, 149.
- Tang, F., Howard, R., and Adkins, J. M.: 1984, Solar Phys. **91**, 75.

- Tarbell, T. D., D. W., Finch, M. L., and Spence, G.: 1981, Proc. SPIE (Suppl.) **265**, 39.
- Teuber, D., Tandberg-Hanssen, E., and Hagyard, M. J.: 1977, Solar Phys. **53**, 97.
- Title, A. M., Pope, T. P., and Andelin, J. P., Jr.: 1974, Appl. Opt. **13**, 2675.
- Title, A. M., and Rosenberg, W. J.: 1979, Appl. Opt. **18**, 3443.
- Title, A. M., and Rosenberg, W. J.: 1981, Opt. Engr. **20**, 915.
- Title, A. M., Tarbell, T. D., Simon, G. W., and the SOUP Team: 1986, Adv. Space Res. **6**(8), 253.
- Topka, K., Moore, R., LaBonte, B. J., and Howard, R.: 1982, Solar Phys. **79**, 231.
- Unno, W.: 1956, Pub. Astron. Soc. Japan **8**, 108.
- Van de Wiele, F.: 1976, "Photodiode Quantum Efficiency," in P. G. Jespers, F. Van de Wiele, and M. H. White (eds.) Solid State Imaging, Noordhoff, Leyden, p. 47.
- Van Hoven, G., Anzer, U., Barbosa, D. D., Birn, J., Cheng, C.-C., Hansen, R. T., Jackson, B. V., Martin, S. F., McIntosh, P. S., Nakagawa, Y., Priest, E. R., Reeves, E. M., Reichmann, E. J., Schmahl, E. J., Smith, J. B., Solodyna, C. V., Thomas, R. J., Uchida, Y., and Walker, A.B.C.: 1980, in P. A. Sturrock (ed.), Solar Flares, Colorado Associated University Press, Boulder, Colorado, p. 17.
- Van Hoven, G., Ma, S. S., and Einaudi, G.: 1981, Astron. Astrophys. **97**, 232.
- Vescelus, F. E., and Antcliffe, G. A.: 1976, Proc. SPIE **78**, 60.
- Wallenhorst, S. G., and Howard, R.: 1982, Solar Phys. **76**, 203.
- Webb, D. F., Davis, J. M., and McIntosh, P. S.: 1984, Solar Phys. **92**, 109.
- Welck, R. E., and Nakagawa, Y.: 1983, NCAR-TN/STR-87.
- White, M. H.: 1976, in P. G. Jespers, F. Van de Wiele, and M. H. White (eds.), Solid State Imaging, Noordhoff, Leyden, p. 485.
- Wight, R.: 1975, in International Conference on the Application of Charge-Coupled Devices, sponsored by Naval Electronics Laboratory Center, p. 121.
- Wilson, R. M., Rabin, D., and Moore, R. L.: 1987, Solar Phys. **111**, 279.

- Wright, J., and Mackay, C. G.: 1981, Proc. SPIE **290**, 160.
- Wu, S. T., Dryer, M., McIntosh, P. S., and Reichmann, E.: 1975, Solar Phys. **44**, 117.
- Wu, S. T., Dryer, M., and Han, S. M.: 1976, Solar Phys. **49**, 187.
- Wu, S. T., Hu, Y. Q., Nakagawa, Y., and Tandberg-Hanssen, E.: 1983a, Astrophys. J. **266**, 866.
- Wu, S. T., Wang, S., Dryer, M., Poland, A. I., Sime, D. G., Wolfson, C. J., Orwig, L. E., and Maxwell, A.: 1983b, Solar Phys. **85**, 351.
- Wu, S. T., Dryer, M., and Han, S. M.: 1983c, Solar Phys. **84**, 395.
- Wu, S. T., Hu, Y. Q., Krall, K. R., Hagyard, M. J., and Smith, J. B., Jr.: 1984, Solar Phys. **90**, 117.
- Wu, S. T., Chang, H. M., and Hagyard, M. J.: 1985a, in M. J. Hagyard (ed.), Measurements of Solar Vector Magnetic Fields, NASA CP-2374, p. 17.
- Wu, S. T., Wang, J. F., and Tandberg-Hanssen, E.: 1985b, in M. R. Kundu and G. D. Holman (eds.), Unstable Current Systems and Plasma Instabilities in Astrophysics, IAU Symp. 107, Reidel, Dordrecht, p. 487.
- Wu, S. T., Dryer, M., and Han, S. M.: 1986a, in T. Watanabe (ed.), Proc. of Symposium on Heliospheric Study, p. 53.
- Wu, S. T., Hu, Y. Q., Nakagawa, Y., and Tandberg-Hanssen, E.: 1986b, Astrophys. J. **306**, 751.
- Wu, S. T., Wang, S., Hu, Y. Q., Michels, D. J., Howard, R. A., Koomen, M. J., and Sheeley, N. R., Jr.: 1987a, submitted to Astrophys. J.
- Wu, S. T., Chang, H. M., and Hagyard, M. J.: 1987b, in preparation.
- Yoder, P. D.: 1986, Opto-Mechanical System Design, Marcel Decker, New York.
- Zirin, H.: 1983, Astrophys. J. **274**, 900.

Sources

Sales information on AMPEREX NXA1011 frame transfer sensor
Amperex Electronic Corp.
Slatersville, RI

Sales information on Fairchild CCD222 interline transfer sensor
Fairchild Corp.
Sunnyvale, CA

Sales information on Videk Megaplug sensor
Videk Megaplug Sensor Corp.
Canandaigua, NY

Sales information on TEKTRONIX CCD imagers
Tektronix Corp.
Beaverton, OR

APPENDIX A

Parameters of Glan-Thompson Polarizers

The following are relationships of parameters that are important in the fabrication of Glan-Thompson polarizers (Bennett and Bennett, 1978).

1. For a constant L/A (length to aperture ratio), the field angle increases with decreasing n , where n is the refractive index of the cement between the prisms.

2. The field angle increases with increasing L/A .

3. $L/A = (1/\tan S)$, where S is the cut angle.

4. One problem with determining the cut angle S for a particular prism is lack of data on the refractive index of cements at different wavelengths. Refractive indices for most cements are given for the wavelength of the Sodium D line.

5. One possible improvement in the Glan-Thompson design is the development of a glass/calcite polarizer. It has the following advantages:

- a. There can be no misalignment of optic axes and thus squirm is eliminated.
- b. The amount of calcite required is reduced; this in turn reduces errors created by using imperfect calcite.

However, this combination creates a different problem: if the elements are cemented together, the thermal expansion of glass is different from calcite.

6. Air spaced polarizers are typically of the Glan-Taylor design. They have the following advantages:

- a. Transmission is 60-100% better than the Glan-Thompson equivalent
- b. Internal reflections are less than 10% of the Glan-Foucault.

The disadvantages are:

- a. They have a limited field of view.
- b. The optical beam will be displaced by a distance that is proportional to the thickness of the calcite (several millimeters).
- c. If half of the prism is glass, the beam can also be deviated (0.5°).

APPENDIX B

Narrow Bandpass Birefringent Filters

W. J. Rosenberg
Lockheed Research Laboratories
Palo Alto, California

1. Introduction

This report describes the characteristics of a birefringent filter which could be incorporated in an orbiting solar vector magnetograph. The design parameters of the filter comply substantially with the functional specification distributed by Russell Chipman (1986). Those specifications of direct interest to the filter design are:

Wavelength Range	5243-5254 Å
Filter Bandpass (FWHM)	120 mÅ
Tunability Resolution	1 mÅ
Clear Aperture	25 mm
Length	not to exceed 33 cm
Angle of Incidence Range	± 24 millirad (in air)

Each of the above parameters are addressed in the subsequent discussion. With the possible exception of the tunability resolution, each of the specifications is quite consistent with a calcite birefringent filter. The tunability resolution question, as will be discussed below, is more a question of precision vs. accuracy and spectral uniformity and stability.

This report is meant to be a discussion of the applicability of a calcite birefringent filter to the solar vector magnetograph. It is not a complete design nor a proposal to build such a filter. We are, however, prepared to produce an explicit filter design and proposal but would need a more complete description of the mission.

2. Standard Lyot Design

The standard birefringent filter is the Lyot design in which the filter consists of a set of crystal plates, each one-half as long as its predecessor (Lyot, 1944; Title and Rosenberg, 1981). Each plate is separated from its neighbors by efficient polarizers having high transmission in the pass direction and high extinction (very low transmission) in the cross direction. The bandwidth, full width at half maximum (FWHM), is determined by the length of the longest element, while the inter-order spacing, the free spectral range (FSR), is determined by the length of the shortest element. Enough elements must be included in order to meet both the FWHM and FSR requirements. For the present application, a FWHM of 120 mÅ requires a longest element length of 60 mm. With seven elements, and consequently a

shortest element length of 0.938 mm, the FSR is nearly 17 Å. Two plots of the resulting filter profile are shown in Figure B.1. The first plot shows a full free spectral range including two filter peaks. A detail of the peak showing the typical Lyot sidelobe structure is shown in the second plot.

The specific plate thicknesses for this seven plate Lyot filter are:

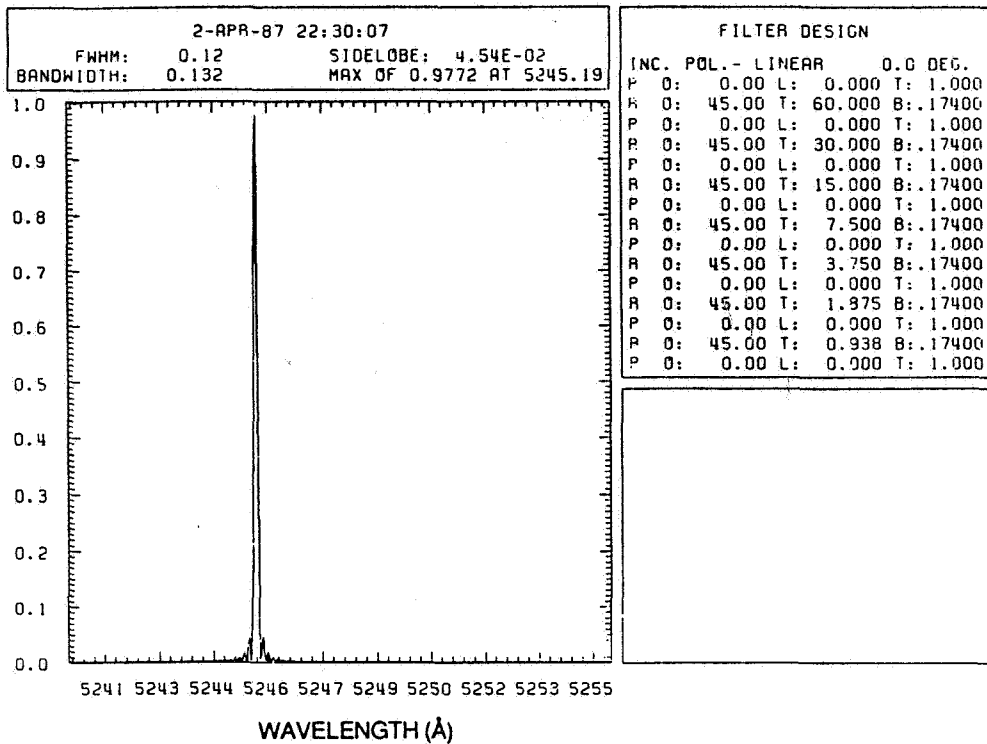
Plate	Thickness (mm)
1	60.
2	30.
3	15.
4	7.5
5	3.75
6	1.875
7	0.938

In practice, it is difficult to make a field widened calcite plate which is less than 1.0 mm thick so the seventh plate must be modified. The best approach is to use a three-halves plate instead of the one-half plate. In this example a three-halves plate is 2.813 mm thick. The resulting change to the transmission profile is shown in Figure B.2. It is apparent that there is almost no change. The side lobe structure is typical of a Lyot design with a maximum side lobe level of about 5% of the main lobe. For the design bandwidth (FWHM) of 120 mÅ the integrated noise bandwidth, or equivalent width, of the filter is 130 mÅ. Of that equivalent width, 95 mÅ, or 73%, falls between the half-maximum points. The ratio of the FWHM passband transmission to that of the wings is, therefore, 2.7.

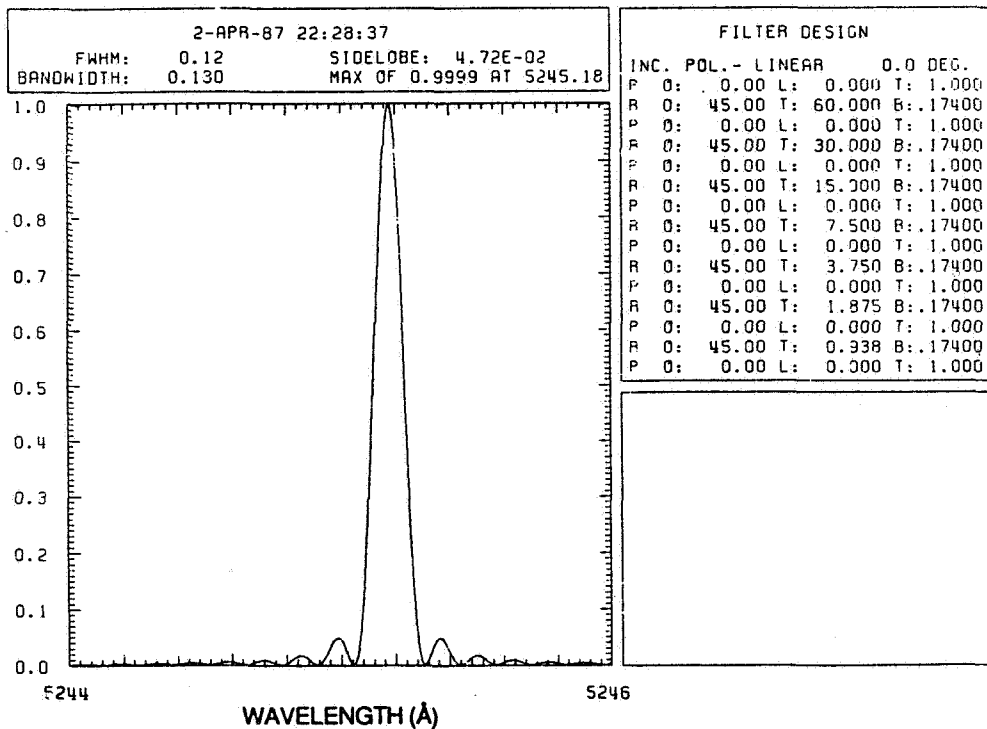
3. Partial Polarizing Configuration

In a manner mathematically equivalent to optical aperture apodization, the response profile of a birefringent filter can be modified to reduce the out-of-band side lobes. While there are a number of techniques which effect this reduction, the most practical is the partial polarizing method (Title and Rosenberg, 1981; Rosenberg, 1986). The calcite elements are separated by leaky, partial, polarizers rather than the high extinction polarizers of the Lyot design. This has two advantages: the side lobes are reduced as desired, and the overall transmission of the filter increases. The transmission increase is due to the polarizers having better transmission in their pass direction as well as leaking light in their extinction direction. The tunable birefringent filter flown on Spacelab 2 is an example of a partial polarizing design.

The spectral profile of a candidate partial polarizing filter is shown in Figure B.3. Again, the upper graph shows a full free spectral range of the spectrum, while the lower graph is a detail of the filter's passband. Notice that the out-of-



a



b

Figure B.1. (a) Transmission profile of seven element Lyot filter across the spectral region of interest. (b) Detail of transmission profile showing passband.

ORIGINAL PAGE IS
OF POOR QUALITY

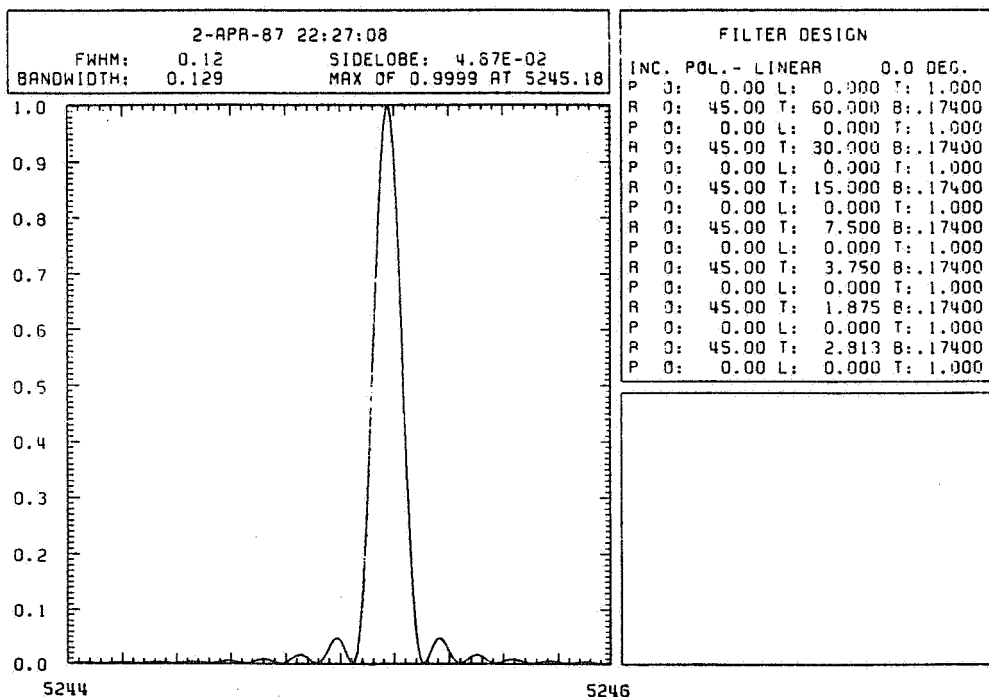
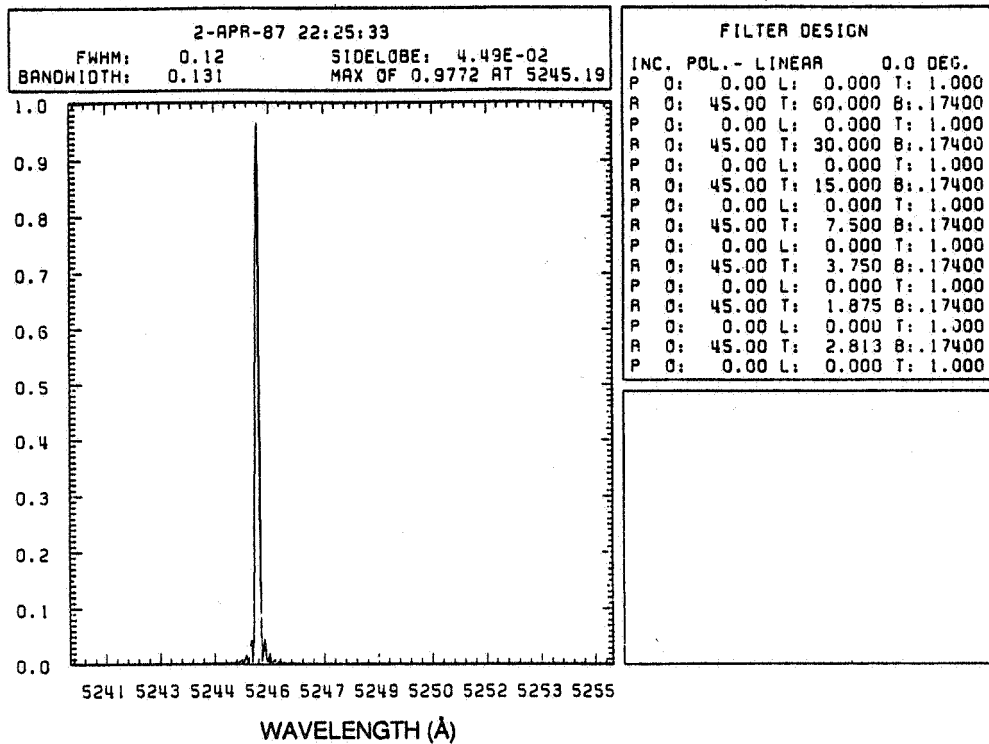


Figure B.2. (a) Transmission profile of modified seven-element filter across the spectral band of interest. (b) Transmission profile detail of the filter passband.

ORIGINAL PAGE IS
OF POOR QUALITY

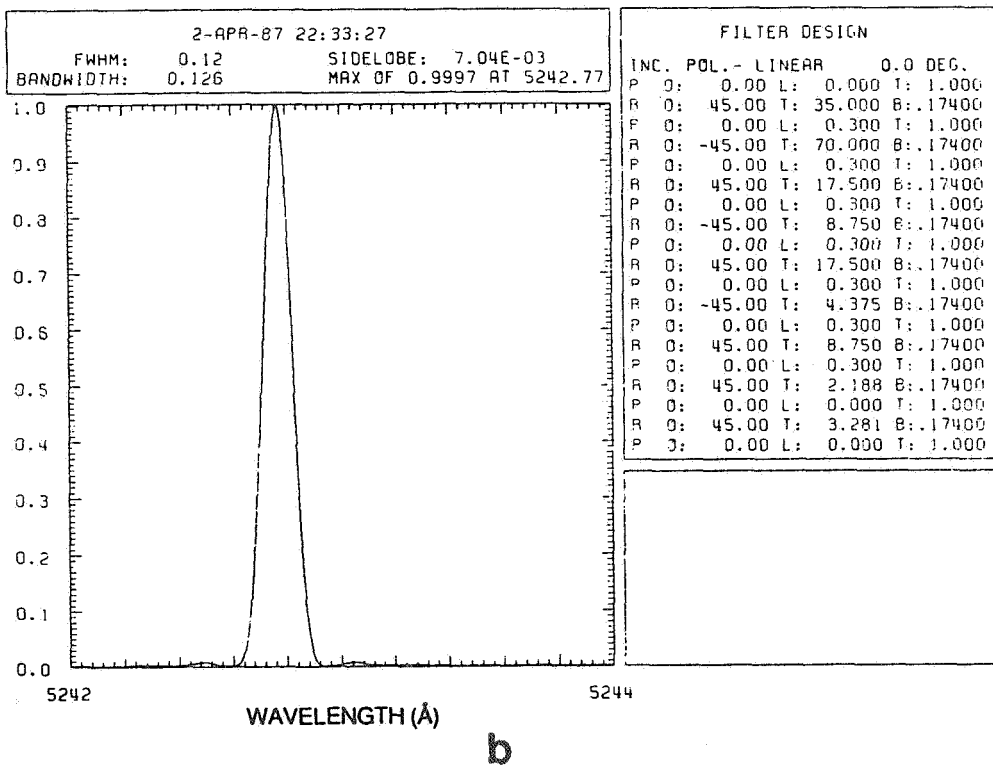
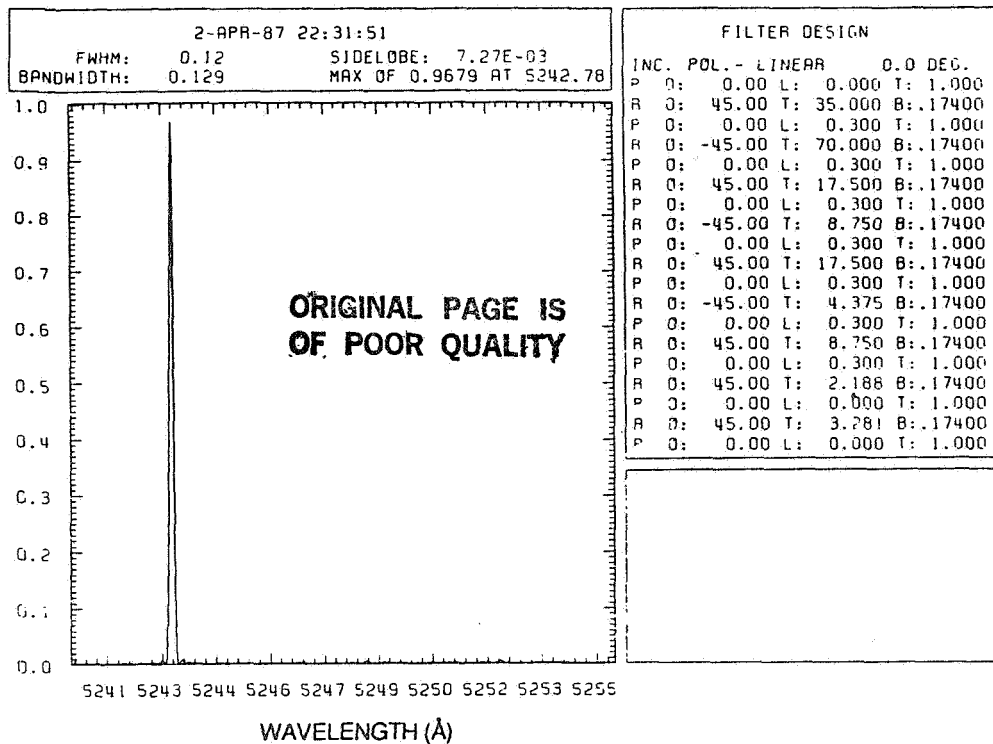


Figure B.3. (a) Transmission profile of the partial polarizing filter design across the spectral band of interest. (b) Transmission profile detail of the partial polarizing filter passband.

band side lobes have been reduced to less than 1% instead of the 5% Lyot side lobe levels. For this partial polarizing design, the FWHM is 120 mÅ and the equivalent width is 126 mÅ. Of that equivalent width, 97 mÅ, or 77%, falls within the half-maximum points. The ratio of the FWHM passband transmission to that of the wings is 3.3.

The specific plate thicknesses for this nine-plate partial polarizing filter are:

Plate	Thickness (mm)
1	35.
2	70.
3	17.5
4	8.75
5	17.5
6	4.375
7	8.75
8	2.188
9	3.281

4. Tuning

One of the most remarkable properties of birefringent filters is the tuning mechanism (Title and Rosenberg, 1981). If a single element is considered, and the exit polarizer is replaced by a quarter-wave retardation plate, then the output of the element is linearly polarized light in which the orientation of the plane of polarization varies with color. A rotatable polarizer following the quarter-wave plate, therefore, becomes a tuner and selects the wavelength which corresponds to its orientation. Intuitively, it can be approached as follows. The output of a birefringent retardation plate whose input is linearly polarized light is elliptically polarized light. Fortunately, the eccentricity of the ellipse varies with wavelength while the orientation of the ellipse axes remains fixed, parallel to the input polarization. The subsequent quarter-wave plate converts elliptical polarization to linear polarization. The orientation of the polarization is dependent on the eccentricity of the ellipse, and therefore on the input wavelength. Tuning is accomplished by a mechanical rotation of a polarizer following the quarter-wave plate.

For Lyot filters with high extinction polarizers the tuning technique works quite well and has been standard practice on birefringent filters for many years. Rotation of the exit polarizer by 180° tunes the element through a free spectral range of the respective element. For the longer, narrower bandwidth elements, the amount of rotation to effect a given wavelength shift becomes larger. The sensitivity to angular errors is constant, independent of the spectral resolution of any given element. It is, therefore, no more difficult to tune a 0.1 Å bandwidth element as a 100 Å one.

Since the orthogonal polarization is absorbed by the rotating exit polarizer in a Lyot configuration, it has not been a concern in most filters. In a partial polarizing system, however, both polarizations are important and must be considered. As it turns out, there is one problem which must be corrected. The orthogonal polarization is one-quarter wave out of phase and will not interfere properly in the subsequent element. This condition is easily corrected by the addition of an achromatic quarter-wave retardation plate oriented parallel to either the transmission or extinction axes of the rotating polarizer. In previous partial polarizing filters using the alternating partial polarizing design this was not necessary because each element has an efficient polarizer at one of its ends. For each element the efficient polarizer end was used for tuning. The all partial polarizing design of a quartz filter built at LMSC in 1986 was the first filter which required the additional quarter-wave plate in its tuning section.

The tuning resolution depends on the angular resolution of the rotation mechanism. A rotation of the polarizer by 180° tunes an element by one free spectral range of that element. For the Lyot design listed above, the longest element is 60 mm long and has a free spectral range of about 260 mÅ. In the partial polarizing design the longest element is 70 mm with a free spectral range of 225 mÅ. In previous filters we have used stepping motors to drive the tuning components. These have always had 48 steps per free spectral range which would correspond to a tuning resolution of 5.5 and 4.7 mÅ for the Lyot and partial polarizing designs, respectively.

There is no reason why greater angular resolution cannot be substituted for the 48 step per revolution motors. In our plans for future filters we are intending to use brushless dc motors with absolute encoders for rotation control. With these the resolution is almost unlimited, being mainly a question of the precision of the encoder. We are specifying 4096 position encoders, not because that accuracy is required, but because those 12-bit encoders are convenient. With such a motor-encoder pair the tuning resolution would be less than 0.1 mÅ. This would be misleading precision, however. With a temperature tolerance of roughly 300 mÅ/°C, a 1.0 mÅ tuning resolution would correspond to a temperature uniformity and stability of only a few millidegrees. Similarly, the wavelength uniformity across the aperture is limited by the thickness uniformity of the calcite crystal element. The wavelength shift is approximately 7 mÅ per micron of thickness variation. With a bandwidth of about 120 mÅ it is extremely difficult to measure thickness variations or wavelength shifts on the order of 1.0 mÅ. While the tuning mechanism could have the required precision, the tuning accuracy would in fact be limited by other factors.

5. Field of View

The feature that made birefringent filters useful to Lyot and subsequent solar physicists is the possibility of relatively large fields of view (Lyot, 1944; Title and Rosenberg, 1979). By constructing a wide field element, the spectral sensitivity to angle of incidence variations can be minimized. A wide field element is constructed by taking a simple crystal element and cutting it into two pieces, each half the length of the original. By placing these two identical plates at a relative orientation of 90° the variation of total retardation with incident angle is minimized. A half-wave retardation plate (or a 90° rotator) separating the two plates is required to really make it work properly. The key to this construction is the azimuthal behavior of individual uni-axial crystal retardation plates. The retardation decreases with incident angle along the optic azimuth and increases on the orthogonal azimuth.

For field widened calcite elements at an incident angle of ± 24 millirad, in air, there is a wavelength shift of $32 \text{ m}\text{\AA}$ relative to the on-axis spectral response. This shift is to the red and is nearly independent of azimuth. Since the intent is to place the filter in a collimated portion of the optical system, this will mean that there will be an apparent red shift of the filter toward the outside of the image plane. The outer pixels will see a $32 \text{ m}\text{\AA}$ redder portion of the spectrum than the on-axis pixels. If the filter were placed at an image plane, in a telecentric system, then there would be no spectral variation over the image. Instead, the apparent bandwidth of the filter would be broadened by the $32 \text{ m}\text{\AA}$, but would be constant over the image. I would certainly consider the effect of changing the optical design to accommodate a telecentric beam at the filter location.

6. Temperature Effects

Calcite birefringent filters drift blue with temperature at a rate of roughly $300 \text{ m}\text{\AA}/^\circ\text{C}$. There are two problems which this causes. First is the change in wavelength of the filter as a whole as its temperature changes, and second is the non-uniformity of the spectral response caused by temperature variations across the aperture of the filter. In our calcite filters we have taken the attitude that since the filter is tunable we could measure the temperature and tune to correct for any changes. With the filter in a sufficiently large aluminum case the thermal inertia makes this quite practical. The aluminum case also serves to minimize the presence of non-uniform temperature distributions at the periphery of the calcite elements. It does not, however, rule out the presence of radial temperature gradients in the crystals. Our observations have shown that this is a surprisingly small effect. We have made measurements of the spectral uniformity across filter apertures and have not found any significant temperature induced nonuniformities. For ambient temperature derivatives of several

degrees C per hour we have not found spectral non-uniformities at the 2 or 3 mÅ sensitivity of our measurements.

One technique which has been suggested for combatting the 300 mÅ per degree drift is to use a compensating crystal in conjunction with the calcite. A promising candidate would be ADP (Ammonium Dihydrogen Phosphate) because of its relatively high retardation variation with temperature. At 5250 ADP has a wavelength shift of almost 5 Å /°C. While this makes it difficult to use as a filter material by itself, it suggests using a relatively short piece to counteract the wavelength shift of calcite. We have not tried this technique ourselves. I have reservations about it because the sensitivity of the ADP might make localized temperature gradients more of a problem. In other words, by trying to fix an already manageable problem we might promote the as yet non-existent problem into real trouble.

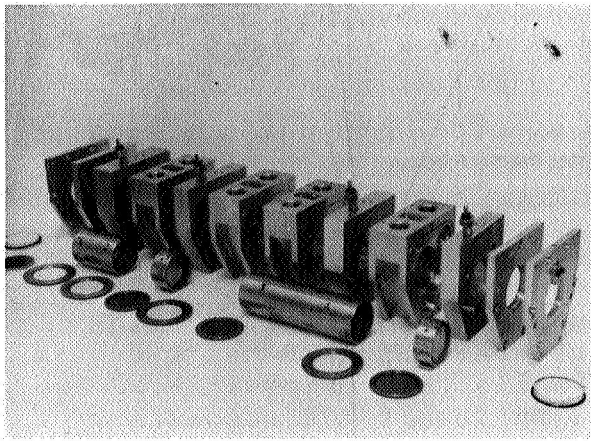
7. Physical Characteristics

The birefringent filters we have built consist of the calcite elements contained in an aluminum package. The elements themselves sit in an oil-filled cavity which provides optical coupling to minimize reflection losses and ghosts, while allowing free rotation of the tuning components. The tuning motors couple to the rotating components through ferrofluidic seals to contain the oil. Previous filters have used 48 position stepper motors with home position sensors to maintain control of the tuning components' rotation. We now believe that a superior choice is a brushless dc motor with high resolution, absolute shaft angle encoders. The Spacelab 2 flight filter package is typical of the physical configuration of a birefringent filter for this application. Its construction is illustrated in Figure B.4.

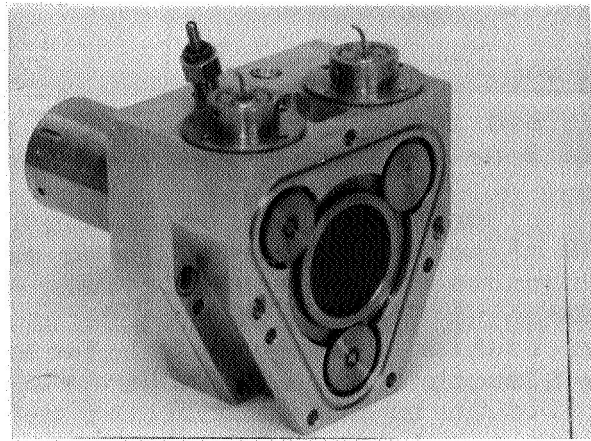
One problem which has been present in the oil-filled tunable filters has been the appearance and subsequent continued presence of small bubbles in the optical path. We believe that better filling techniques, currently being employed in the Spacelab 2 filter refurbishment, will mitigate against the formation of bubbles. A packaging redesign which minimizes small passages and closely spaced rotating components should facilitate the removal of any bubbles which do form.

The question of clear aperture depends almost entirely on calcite availability. At the present time there appears to be a supply of calcite crystals consistent with a 25-mm clear aperture. There is the possibility, though not yet established, that larger aperture crystals might be available. We are thinking in terms of 50 mm to possibly 75 mm clear apertures. This possibility is being pursued independently of this application.

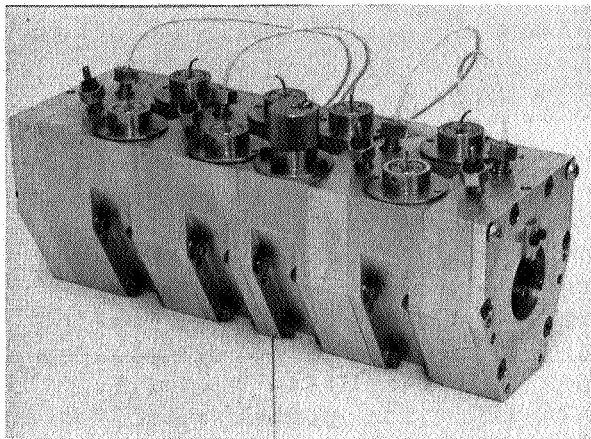
The length of the filter depends on the choice of filter designs. For the Lyot design there is a total length of calcite of 121 mm. The seven elements require approximately 54 mm of



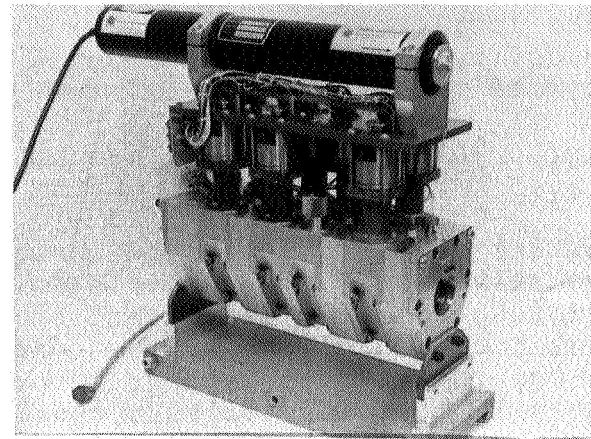
A



B



C



D

Figure B.4. Illustration of the assembly complexity of a flight-qualified tunable filter showing: (a) complete kit of optical and case components; (b) assembly detail of single case section including calcite housing holding the two longest elements; (c) completed assembly of optical and case components including rotary penetration joints for tuning motor coupling; (d) complete filter including tuning motors and calibration laser.

ORIGINAL PAGE IS
OF POOR QUALITY

wave plates, polarizers, and clearance. This implies a total optical length of about 175 mm, though it could be slightly more depending on the exact element packaging technique. The partial polarizing design uses about 170 mm of calcite, and requires about 70 mm of overhead for its nine elements. This implies a length estimate of 240 mm.

8. Transmission

The transmission of the Spacelab 2 flight filter at 5309 Å is 14% in polarized light. This is representative of what would be expected in this application. Since the filter need only tune over a limited range around 5250 Å it should have somewhat better transmission. The polarizers can be optimized for the blue-green and the various wave plates can be simple monochromatic retarders instead of the achromatic laminations required for a universal filter. Since estimates of transmission often assume as much as 1% loss per layer of oriented plastic film, the use of single layer monochromatic retarders has a significant effect in a seven- or nine-element filter.

The birefringent filter would only be blocked over the 12 Å region around 5250 Å. In order to isolate that band a narrow passband, multi-layer dielectric, interference filter will be required. The transmission spectrum of a typical filter is shown in Figure B.5.

DATE: 03:26:87 OPERATOR: _____
CELL: _____ TEMPERATURE: 24° C
HUMIDITY: _____

SAMPLE: 5250 Å FLT
SOLVENT: B/W 5A-7A
CONC: METHOD 55 USED

DATA MODE : (1) %T
BAND WIDTH : 0.10 nm
TIME CONST : 0.4 sec
λ SET : 527.0 ~ 522.0 nm
λ SCALE : 0.5 nm/cm
SCAN SPEED : 4 nm/min
%T ABS SCALE : 0.0 ~ 50.0 %T
CYCLE NO : 1

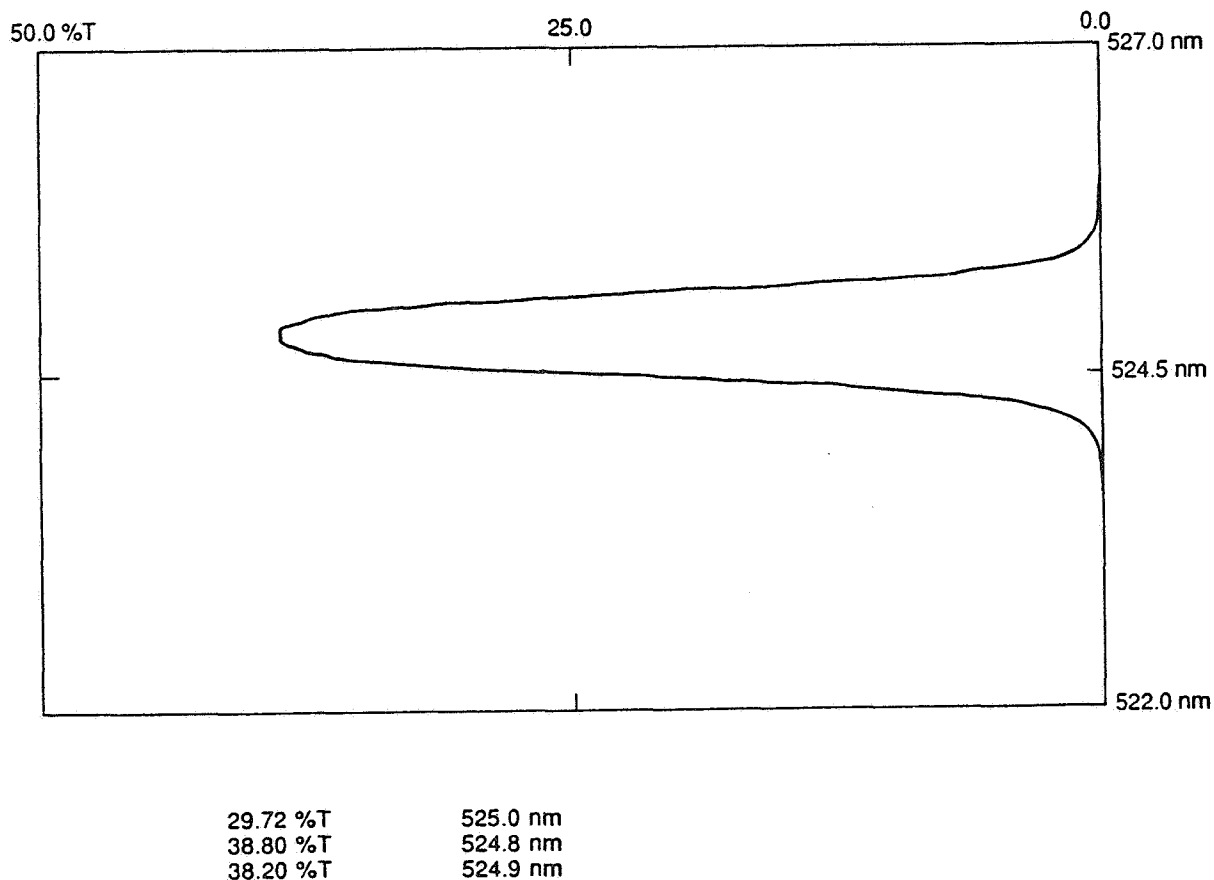


Figure B.5. Spectrophotometric measurement of multi-layer dielectric blocking filter for the 525 nm spectral region. This is a flight-qualified, three-period, interference filter.

REFERENCES

Chipman, R.: 1986, Tunable Filter Optical Specification, Space Flight Solar Vector Magnetograph Design Project, NASA Marshall Space Flight Center.

Lyot, B.: 1944, Ann. Astrophys. 7, 31.

Title, A. M., and W. J. Rosenberg: 1981, Opt. Engr. 20, 915.

Rosenberg, W. J.: 1986, Compact Photometric and Imaging Spectroscopic System, Final Report NASW-3662, NASA Headquarters (LMSC-F108007).

Title, A. M., and W. J. Rosenberg: 1979, Appl. Opt. 18, 3443.

APPENDIX C

Use of a Fabry-Perot Filter in a Space Flight Solar Vector Magnetograph

D. M. Rust, T. Appourchaux, and T. J. Harris

The Johns Hopkins University
Applied Physics Laboratory
Laurel, MD 20707

ABSTRACT

In response to a request from the NASA Marshall Space Flight Center's Space Science Laboratory, we have made a preliminary study of how a recently developed filter based on a solid Fabry-Perot etalon can be used in a solar vector magnetograph. The Fabry-Perot filter is electrically tunable, compact and durable, and apparently suitable for use on a spacecraft. Principal topics covered in this report are spectral filtering characteristics, performance in a candidate optical system, and ability of the filter to operate in space.

1. Introduction

Solar flares and their effects on man's activities in space are unpredictable at present, in large part because of inadequate knowledge of the solar magnetic fields and the physical processes induced by their presence in the solar atmosphere. Although substantial progress toward understanding solar magnetism has been made over the past two decades, high-resolution observations from space will be required to understand the subtle field developments that seem to be associated with solar activity (Rust and Bridges, 1975; Martin et al., 1985). Atmospheric blurring of solar images usually makes long time series of high-precision, high-resolution magnetograms impossible to obtain with ground-based telescopes. Therefore, we have explored design criteria for a spaceborne solar magnetograph, specifically, one based on use of a solid Fabry-Perot etalon for spectral analysis of atomic absorption lines in the solar spectrum.

The basic principle of operation of a magnetograph can be understood by considering a Fraunhofer (atomic absorption) line profile convolved with the profile of a narrow-band filter. This is treated in detail in section C.II (Spectral Filtering Characteristics), but, in brief, a bandpass filter tuned through a Fraunhofer line profile is used to measure the velocity and magnetic field at the imaged, radiating region of the solar surface, by relating the spectral line properties to the Doppler and Zeeman effects, respectively, in the radiating region. To measure the full vector magnetic field, the line must be sampled at several points on its profile and all four Stokes polarization parameters measured at each point (Hagyard, 1985).

The key requirements for a vector magnetograph are fine spectral resolution and ultra-low instrumental polarization. The spectrographic requirement can be met by especially designing an optical system around a high throughput, tunable, solid Fabry-Perot (FP) etalon such as the one recently developed for solar seismology by Rust et al. (1986a,c). We will concentrate on this aspect of the magnetograph and defer discussion of the polarization properties of the optical system to a later article.

A solid etalon with a high index substrate requires a much smaller aperture than does an air-spaced etalon to attain the same Lagrangian invariant and filter half-width. This and other optical characteristics of the FP-based magnetograph are treated in more detail in section C.III (Etalon Performance in an Imaging System).

The space environment places special restrictions on the materials and devices to be incorporated in a magnetograph. Although the space worthiness of fused quartz FP etalons was established during the Skylab mission, questions may still be raised concerning the ability of a crystalline etalon to survive launch vibration, cosmic radiation and years in a vacuum. This topic is considered in section C.IV (Survival in Space).

2. Spectral Filtering Characteristics

2.1 Etalon Design Considerations

The performance of a solid FP etalon in an optical system depends on the index of refraction of the substrate, the free spectral range, the finesse, the diameter of the clear aperture, and placement relative to the pupil and image planes. The etalons under consideration here are constructed of lithium niobate, which is a highly transparent crystalline material whose index of refraction changes in proportion to the applied voltage. The advantage of using lithium niobate crystal is that the FP passband can be tuned by the application of voltage to the crystal faces (Gunning, 1982). For lithium niobate, a voltage of ± 250 V will shift the etalon's passband by ± 0.1 Å.

In an FP, the higher the refractive index of the spacer, the less the passband shifts and broadens for off-axis rays. The wavelength shift is given by

$$\Delta\lambda/\lambda = \phi^2/2n^2, \quad (1)$$

where ϕ is the angle between the incoming ray and the normal to the filter and n is the index of refraction of the spacer. Thus, the acceptance angle of the lithium niobate etalon, for which $n = 2.3$, is 5.3 X that for an air-spaced etalon. Another way to express this advantage is to note that for the same spatial and spectral resolution, an air-spaced etalon must be 5 times the diameter of a lithium niobate etalon.

The free spectral range FSR of an etalon of thickness d and index n is given by

$$\text{FSR} = \frac{\lambda^2}{2nd}, \quad (2)$$

and the full width at half maximum (FWHM) of the passband is FSR/F , where F is the effective finesse, which depends upon the flatness, smoothness, and reflectivity of the coated surfaces. Figure C.1 shows the transmission characteristics of a lithium niobate etalon made for us at the Commonwealth Scientific and Industrial Research Organization (Australia) Division of Applied Physics. The thickness of the etalon was 0.22 mm and the wavelength of the light was $\sim 5700 \text{ \AA}$.

Experience with a lithium niobate etalon (Rust et al., 1986b) has established that a 50-mm etalon with $\text{FWHM} = 175 \text{ m\AA}$ is suitable for two-point observations of a Fraunhofer line profile in a low-resolution imaging system. In a study of possible etalons for the high-resolution vector magnetograph, we have arrived at the following etalon specifications:

Filter passband (FWHM)	120 mÅ (at 5250 Å)
Finesse	22
Free spectral range (FSR)	2.6 Å (at 5250 Å)
Filter transmission	53%
Clear aperture	62 mm
Thickness	0.22 mm
Angle of incidence range in air	$\pm 6.3 \text{ mrad}$

The operating wavelength and the FWHM were chosen on the basis of properties of the solar spectral lines and to maximize the signal-to-noise ratio (see section C.III). The FSR depends on the thickness of the etalon, as Figure C.2 shows. The FSR should be as large as possible to ease the requirements on the auxiliary filters that will be needed to block all FP orders but one. However, it may be difficult to make a lithium niobate spacer much thinner than the 0.22 mm ones already in hand. Thus, we take $d = 0.22 \text{ mm}$ and $\text{FSR} = 2.6 \text{ Å}$ (ref. Fig. C.2). Then, the required effective finesse (FSR/FWHM) is 22.

Several factors contribute to the effective finesse, but for substrates polished to better than $\lambda/200$, the effective finesse is determined by the reflectivity of the coatings:

$$F = \pi / \ln R, \quad (3)$$

so, coatings with a reflectivity of 87% are called for.

The maximum transmittance T_{max} of an etalon is determined primarily by the reflectance R of the coating and its absorptance A :

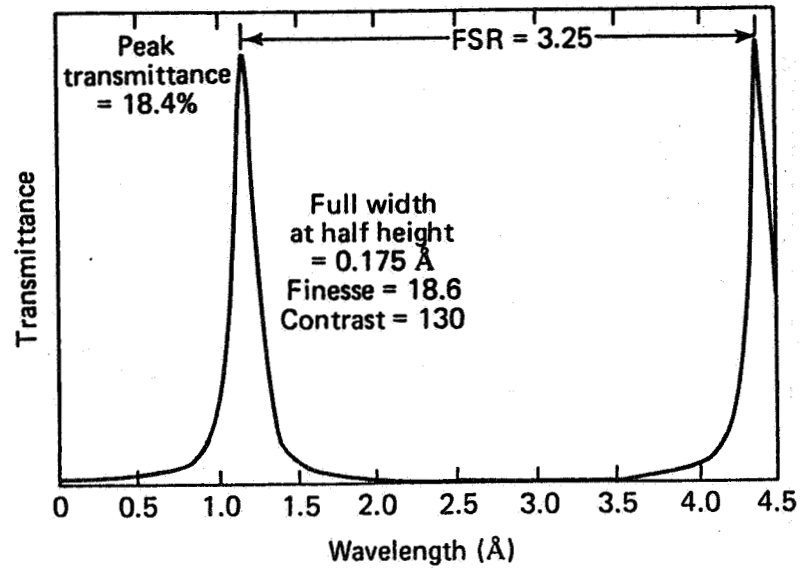


Fig. C.1. Measured transmittance of a Lithium-Niobate FP. Wavelength scale is relative. Measurements made near 5700Å.

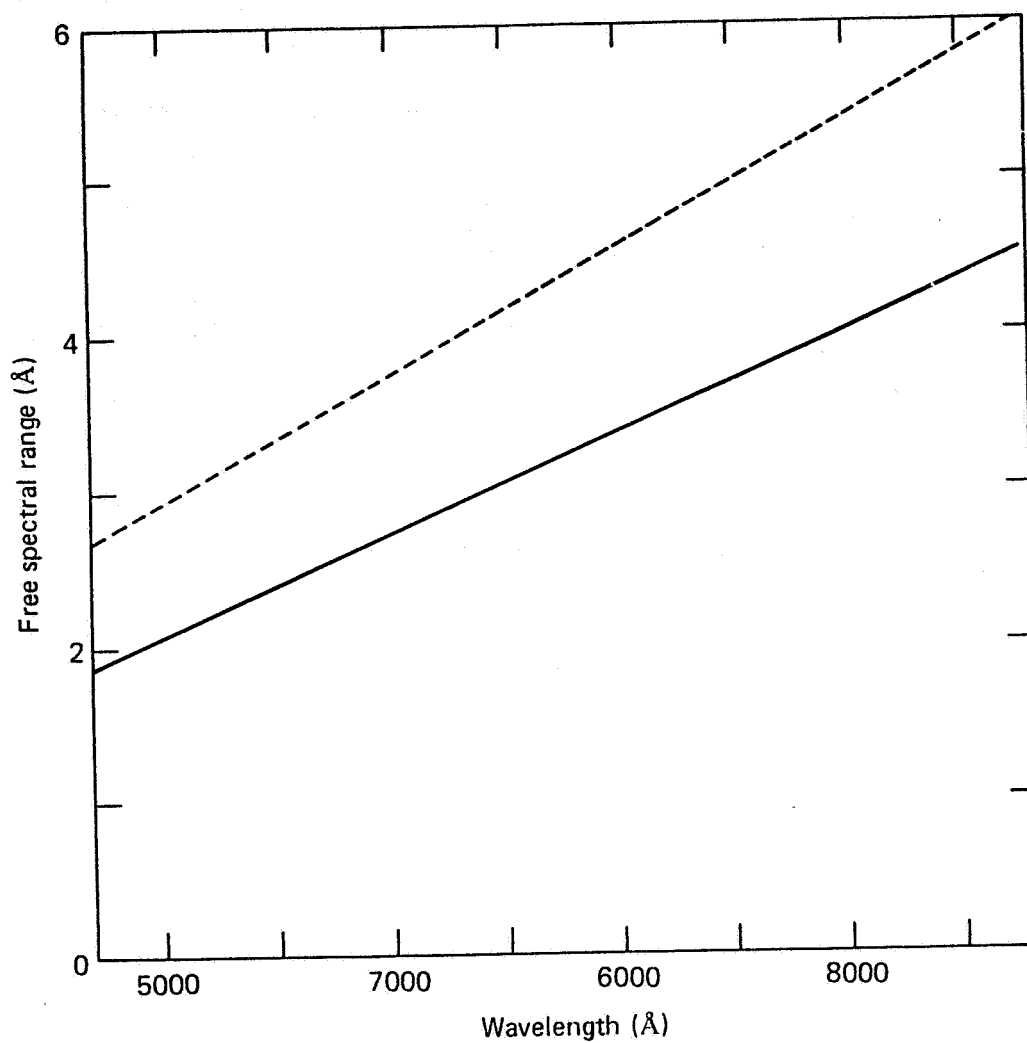


Fig. C.2. Free spectral range as a function of wavelength. Solid curve: Etalon thickness: 250 μm ; dashed curve: 175 μm .

$$T_{\max} = (1 - A/(1-R))^2 \quad . \quad (4)$$

Silver is one practical choice for the coating material. For a 410 Å layer of silver, $R = 0.87$ and $A = 0.035$ in the 5000-6000 Å range, hence T_{\max} is 53%. Multilayers of dielectric material (e.g., SiO_2 and TiO_2) are more durable than silver and one can achieve higher reflectance and thus greater finesse, although possibly at the expense of somewhat lower T_{\max} .

The clear aperture of 62 mm is near the largest practical size because lithium niobate is available in wafers only up to 75 mm diameter, and it is not always possible to achieve the degree of flatness and parallelism near the edges of the etalon that is required to give a finesse >20 there.

The angle of incidence range is found by dividing the system Lagrangian invariant of 0.195 mm radian (section C.III) by the semi-diameter of the etalon. The effect of a range of angles of incidence on the filtering properties depends on the optical system. In the telecentric system discussed below, the rays from a point in the image plane traverse the FP at the same angle, but this angle varies by ± 6.3 milliradians across the image. Therefore, for points at the extreme edges of the field of view, the filter passband will be offset by 20 mÅ from that seen by on-axis rays. There will be no additional broadening of the filter profile, however.

2.2 Tuning

The tuning requirements of a solar magnetograph are modest. Only Doppler shifts due to orbital motion of the spacecraft must be compensated for. These will be less than 0.2 Å. To shift the passband from one wing of the spectral line to the other also requires ~ 0.2 Å tunability. Thus, ± 1000 V will provide adequate control over the passband. The tuning resolution or step size in wavelength can be of any magnitude from 0.4 Å down to 4 μÅ corresponding to voltage steps of 1000 V and 10 mV, respectively.

If the allowed voltage range is restricted to ± 1000 V, then only 30% ($0.8/2.6$) of the spectrum is accessible, and the lithium niobate FP cannot be considered a "universal filter." However, the breakdown voltage of LiNbO_3 is much greater than 1000 V, probably over 4000 V (Gunning, 1982). After further testing, we expect that the operational voltages can be increased to 3000 V, at which point the filter will be able to tune to any wavelength, if provided with adequate blocking filters.

The FP may also be tuned by changing the temperature, if time is not critical. At 5250 Å, a temperature change of 20 °K produces a wavelength shift of 0.76 Å. During solar

observations, the etalon's temperature must be held constant to avoid distorting the magnetic field measurements. The temperature controller for our existing etalon holds the temperature constant to ± 10 mK. The induced uncertainty of 0.38 mÅ in the passband position is of no importance in solar magnetography. However, during an extended space mission it is important to remeasure the central wavelength, voltage tuning parameter, and filter profile from time to time. For the lithium niobate FP, no special monitoring system is required for this because all critical filter parameters can be derived by tuning through a specified portion of the solar spectrum, which provides an adequate standard in itself.

The solar spectrum in one region of interest for magnetography is shown in Figure C.3. Lines marked C and E are separated by 2.65 Å. With very skillful polishing techniques, an etalon may be worked to precisely that thickness which corresponds to a FSR = 2.65 Å. Thus, both lines may be sampled simultaneously by adjacent FP orders. This would allow the use of a 5-Å blocker to reject the unwanted orders rather than a 2.5-Å blocker, which would have a lower peak transmission. The simultaneous use of two lines would also improve the signal-to-noise ratio by $\sqrt{2}$.

2.3 Line Spread Function

When used on solar spectral lines, the line spread function T of an FP is defined by the Airy function:

$$T(v_R) = \frac{1}{1 + \frac{4F^2}{\pi^2} \sin^2\left(\frac{\pi K v_R}{F_R}\right)} \quad (5)$$

with F = finesse

K = order of the FP = $2nd/\lambda$

d = thickness (plate separation)

n = index of refraction

F_R = "finesse" of the solar line = FKx ; x is the FWHM of the FP relative to the FWHM of the line
 $x = \Delta v_{FP}/\Delta v_0$.

v_R = offset, expressed as velocity relative to the width of the line (Δv_0).

This function can be approximated accurately if:

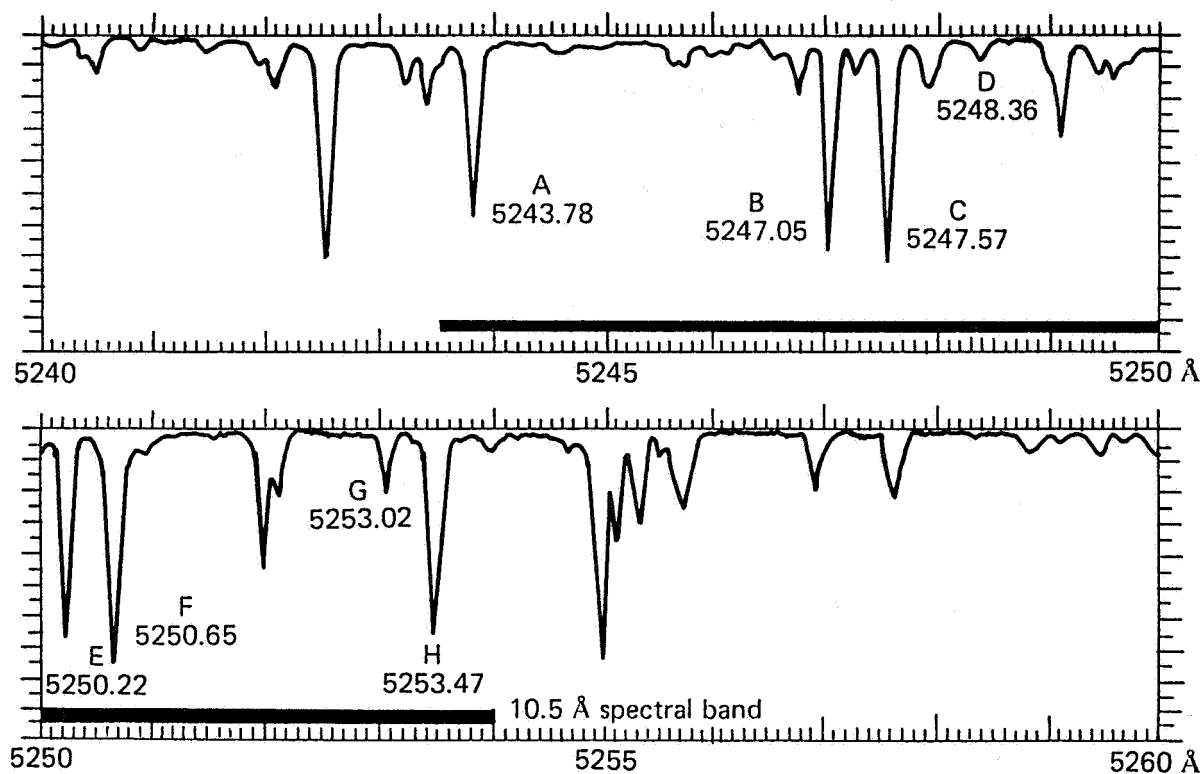


Fig. C.3. Solar spectrum used by the magnetograph. The thick, dark bar indicates the working range. The wavelengths of the lines are indicated.

$$\pi K \frac{v_R}{F_R} \ll 1$$

as

$$T(v_R) = \frac{1}{1 + 4 \left(\frac{v_R}{x} \right)^2}, \quad (6)$$

i.e., the profile is Lorentzian.

2.4 Signal-to-Noise Ratio With One Line

The longitudinal component of the magnetic field is derived from two measurements: $\Delta\lambda^+$ and $\Delta\lambda^-$ which are, respectively, the wavelength shifts for the right-circularly and left-circularly polarized components of the line.

$$\Delta\lambda^+ \sim \left(\frac{D}{S} \right)^+ \equiv \frac{I_r^+ - I_b^+}{I_r^+ + I_b^+} \quad \Delta\lambda^- \sim \left(\frac{D}{S} \right)^- \equiv \frac{I_r^- - I_b^-}{I_r^- + I_b^-}, \quad (7)$$

where I_r and I_b represent the intensity of light detected when the FP is tuned to the red and blue wings, respectively, at the spectral line. The longitudinal magnetic field is proportional to $1/2 (\Delta\lambda^+ - \Delta\lambda^-)$.

We can approximate both measurements in the linear case

$$(I_r^+ \sim I_b^+, I_r^- \sim I_b^-):$$

$$\left(\frac{D}{S} \right)^\pm = \pm \frac{2\delta v_B}{I(y/2)} \frac{\partial I}{\partial y}, \quad (8)$$

where δv_B is the velocity equivalent of a shift due to a longitudinal magnetic field B and $I(y/2)$ is the intensity measured with an offset $y/2$ from the center of line. (y is non-dimensional and relative to Δv_0)

$$\delta v_B = (1.2)^{-1} g B / \Delta v_0 \quad \begin{array}{l} B \text{ in gauss} \\ g = \text{Landé factor} \\ \delta v_B \text{ non-dimensional} \end{array} \quad (9)$$

The uncertainty of the measurement of $\Delta\lambda^+ - \Delta\lambda^-$ is evaluated as follows: the uncertainty in $\left(\frac{D}{S} \right)^+ - \left(\frac{D}{S} \right)^-$ is calculated, taking

into account photon noise, and the linear case formula is used to find the equivalent rms magnetic field (the least magnetic field detectable).

Thus:

$$\delta v_B = \frac{I(y/2)}{4} \left(\frac{\partial I}{\partial y} \right)^{-1} \left[\left(\frac{D}{S} \right)^+ - \left(\frac{D}{S} \right)^- \right] \quad (10)$$

and

$$\delta v_B^{\text{rms}} = \frac{I(y/2)}{4} \left(\frac{\partial I}{\partial y} \right)^{-1} \Delta \left(\left(\frac{D}{S} \right)^+ - \left(\frac{D}{S} \right)^- \right), \quad (11)$$

where

$$\Delta \left(\left(\frac{D}{S} \right)^+ - \left(\frac{D}{S} \right)^- \right) = 2 \frac{1}{\sqrt{2} I(y/2)}. \quad (12)$$

The last expression, for the standard deviation, assumes that photon noise is the principal source of uncertainty.

Using (9), (11), and (12) we find:

$$B_{\parallel}^{\text{rms}} = \frac{1.2}{g} \frac{\Delta v_0}{\sqrt{N}} Q(x, y), \quad (13)$$

where $Q(x, y)$ = quality factor for the spectral line. It is defined as:

$$Q(x, y) = \frac{1}{2\sqrt{2}} \left(\frac{\partial I_R}{\partial y} \right)^{-1} \sqrt{I_R(y/2)} \quad (\text{see Figure C.4}).$$

In the formula (13) N is the number of photons collected by an FP with a rectangular profile of width Δv_0 in unpolarized light.

$$N = U N_0 \Delta v_0 \frac{\lambda_0}{c} T_0 t, \quad \text{where}$$

U = étendue in $\text{mm}^2 \text{ster}$

N_0 = solar flux in photons $\text{ster}^{-1} \text{mm}^{-2} \text{\AA}^{-1} \text{s}^{-1}$

Δv_0 = width of the line in m s^{-1}

λ_0 = wavelength in \AA

c = speed of light in m s^{-1}

T_0 = optical transmission

t = integration time in seconds .

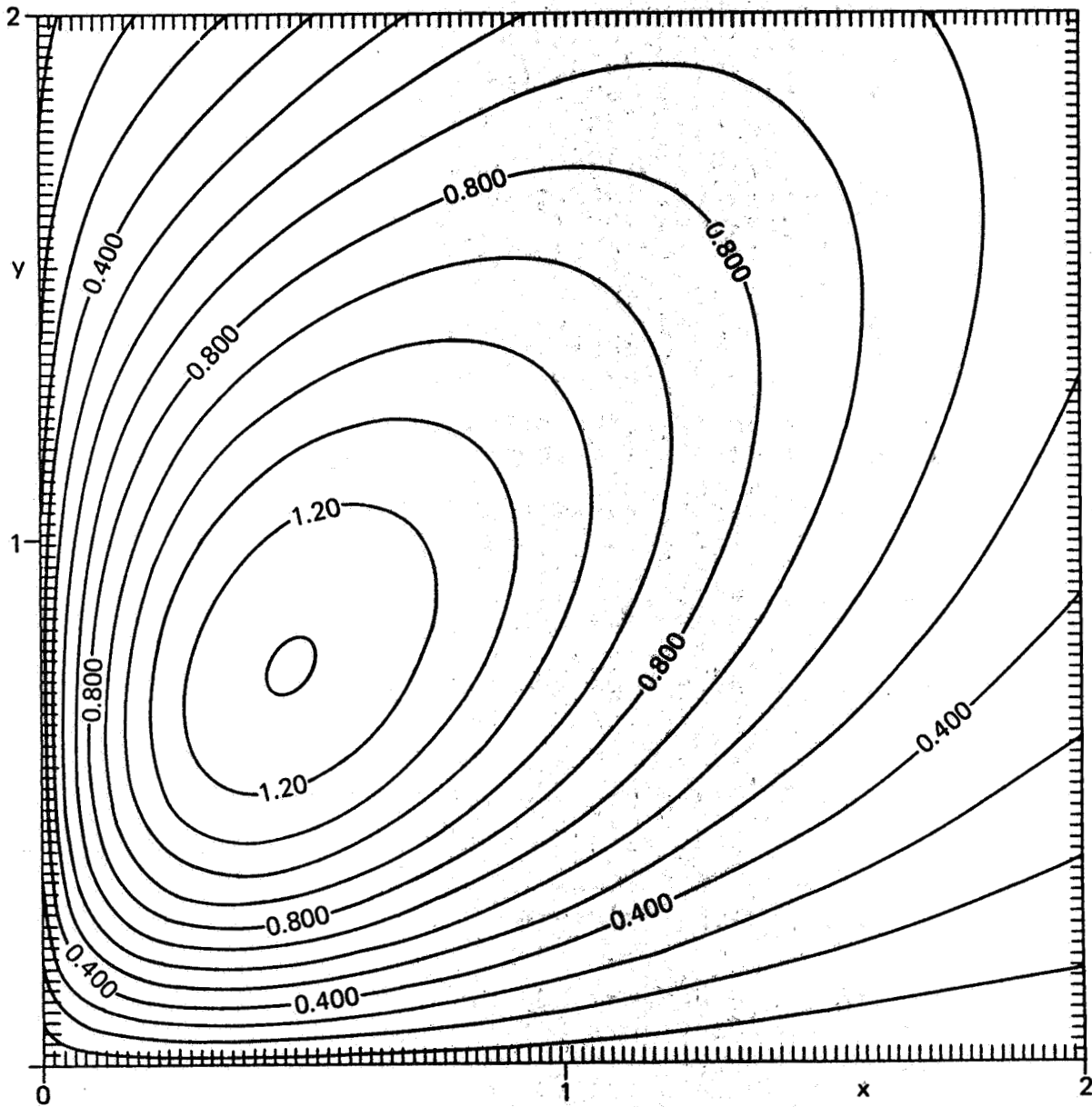


Fig. C.4. Isocontours of the Q^{-1} factor as a function of x and y for a line depth of $P = 0.8$.

$I_R (= I/N)$ is the intensity of the Fraunhofer line convolved with the Lorentzian (or Airy) profile of the FP relative to that produced by a rectangular profile.

$$I_R(y/2) = \int_{-\infty}^{+\infty} F R(v_R) T(v_R - y/2) dv_R . \quad (14)$$

$F R(v_R)$ is the normalized profile of the Fraunhofer line.

2.5 Signal-to-Noise Ratio With Two Lines

With two lines (1 and 2) separated by exactly one FSR of the FP, one may show that

$$\Delta((\frac{D}{S})^+ - (\frac{D}{S})^-) = 2 \{(\frac{D}{S})^2\}^{1/2} = 2 \frac{1}{\sqrt{2} (I_R^1(y/2) + I_R^2(y/2))^{1/2}} . \quad (15)$$

Thus as in eq. 14 we find

$$B_{2 \text{ lines}}^{\text{rms}} = 1.2 \frac{1}{\sqrt{N}} \frac{1}{2\sqrt{2}} \left[\frac{g_1}{\Delta v_0^1} \frac{\partial I_R^1}{\partial y} + \frac{g_2}{\Delta v_0^2} \frac{\partial I_R^2}{\partial y} \right]^{-1} \frac{1}{(I_R^1(y/2) + I_R^2(y/2))^{-1/2}} . \quad (16)$$

If the two profiles are the same $I_R^1(v_R) = I_R^2(v_R)$, $\Delta v_0^1 = \Delta v_0^2$, and $g_1 = g_2$.

In this case $B_{2 \text{ lines}}^{\text{rms}} = B_{1 \text{ line}}^{\text{rms}} / \sqrt{2}$.

2.6 Transmission Goals and Expected Results

We use the relation

$$B^{\text{rms}} = \frac{1.2}{g} \frac{\Delta v_0}{\sqrt{N}} Q(x, y) \quad (17)$$

to determine the rms minimum detectable magnetic field. For the optical system described below,

$$U = 0.194 \text{ mm}^2 \text{ ster}$$

$$N_0 = 8.14 \times 10^{15} \text{ photons ster}^{-1} \text{ mm}^{-2} \text{ A}^{-1} \text{ s}^{-1}$$

$$\Delta v_0 = 8.6 \times 10^3 \text{ m s}^{-1} (\Delta \lambda_0 \approx 0.15 \text{ A})$$

$$\lambda_0 = 5250 \text{ \AA}$$

$$c = 3 \times 10^8 \text{ m s}^{-1}$$

$$T_0 = 0.28 \times 0.53 \times 0.5 \times 0.1 = 7.4 \times 10^{-3}$$

(blocker x FP x DQE x transmission of all lens, mirrors, etc.)

$$Q^{-1} \sim 0.1, 0.5 \quad (\text{see Figure C.5})$$

$$g \approx 1.5$$

$$N(e^-/s) = 9.1 \times 10^{12} \text{ U } e^-/s . \quad (18)$$

This is the number of electrons per second generated as though a single detector sampled the whole field of view. In the high-resolution magnetograph, the throughput must be divided by the number of picture elements (pixels) in the CCD. We take this to be $(2048)^2$, so with the value of U given above, the number of electrons generated per pixel is

$$N_{px} \approx 4.4 \times 10^5 \text{ e}^-/s . \quad (19)$$

The 3/4 saturation level of a CCD that saturates at 750,000 e^- is reached in 1.25 sec.

Assuming this working level ($5.65 \times 10^5 \text{ e}^-$), and $g \approx 1.5$, Q^{-1} , $\approx 0.1, 0.5$,

$$B^{rms} \approx 9 \text{ Q } (e^- \text{ level fixed}) \quad (20)$$

or, generally

$$B^{rms} \approx 10 \text{ Q } t^{-1/2} \quad (21)$$

for quarter-arc sec pixels.

Thus, the rms minimum detectable longitudinal field (for e^- level fixed) is ~ 20 to 100 gauss, for 1-sec integration time, depending on Q mainly. This is given for a single pixel. The more often used resolution element of ~ 2.5 arc sec would allow the same sensitivity in 0.01 sec. Uncertainties in measuring the full vector field are much greater (see the Addendum following section C.V).

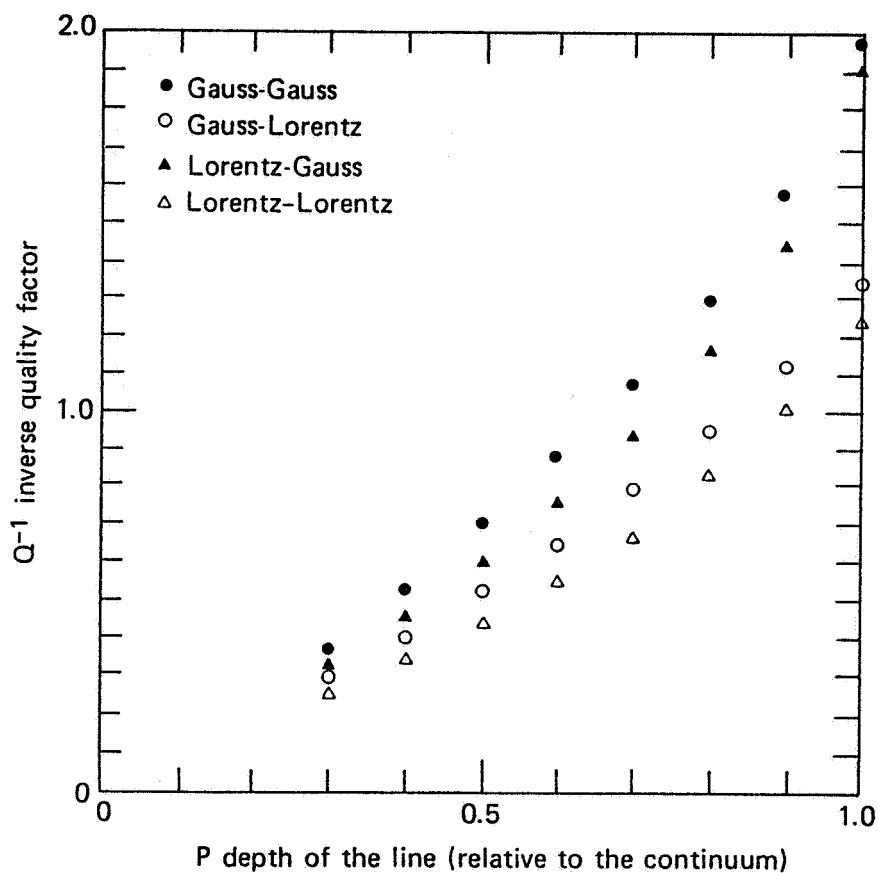


Fig. C.5. Variation of the maximum Q^{-1} factor as a function of the depth of the line for different line, filter profile convolutions.

2.7 Effect of Pointing Errors

The variation of the apparent mean velocity as a function of guider offset error is given by:

$$\delta v = v_E \delta X_O. \quad (22)$$

δX_O is the angular guiding error, relative to the radius of the Sun. Only the component along the solar equator is of concern here because it is the rotation velocity which gives rise to the error δv . v_E is the solar equatorial velocity ($v_E \approx 2 \text{ km s}^{-1}$). We neglect differential rotation and keep only the effect of the rigid body rotation. The variation δv is independent of the Fabry-Perot tilt because its field of view does not move with the guiding. Suppose the guiding error to be ≤ 1 pixel in both directions ($\sim 0''.25$). Then, $\delta v \approx 0.5 \text{ m s}^{-1}/\text{pixel}$. This should have no effect on the measurements since $\delta v \ll \Delta v_0$.

2.8 Velocity Variation in the Field of View

To make reliable magnetic field measurements, the velocity variations across the field of view must be less than the width of the line Δv_0 . In a first step, we put an upper limit to this variation, and in a second step we give a solution to make the variation as small as possible.

To a first approximation we assume that the Sun is rotating like a rigid body. The apparent velocity on any point of the Sun is given by:

$$v(x,y) = v_E X + \alpha(X^2 + Y^2) \quad (23)$$

$$\alpha = \gamma^2 \frac{c \theta^2}{2n^2}$$

c speed of light
 θ radius of the Sun
 n index of refraction of the lithium niobate
 γ magnification of the Sun at the FP. $\gamma = \frac{156}{131.2} = 5$
 $\alpha \approx 15.36 \text{ km s}^{-1}$

The isotachs are circles nearly centered on the origin

($\frac{v_E}{\alpha} \sim 0.13$). The variation of velocity ΔV for a rectangular

field whose corners are located on the half-plane $X > 0$ is given by:

$$\Delta V = v_E \Delta X_O + 2\alpha (X_O \Delta X_O + Y_O \Delta Y_O). \quad (24)$$

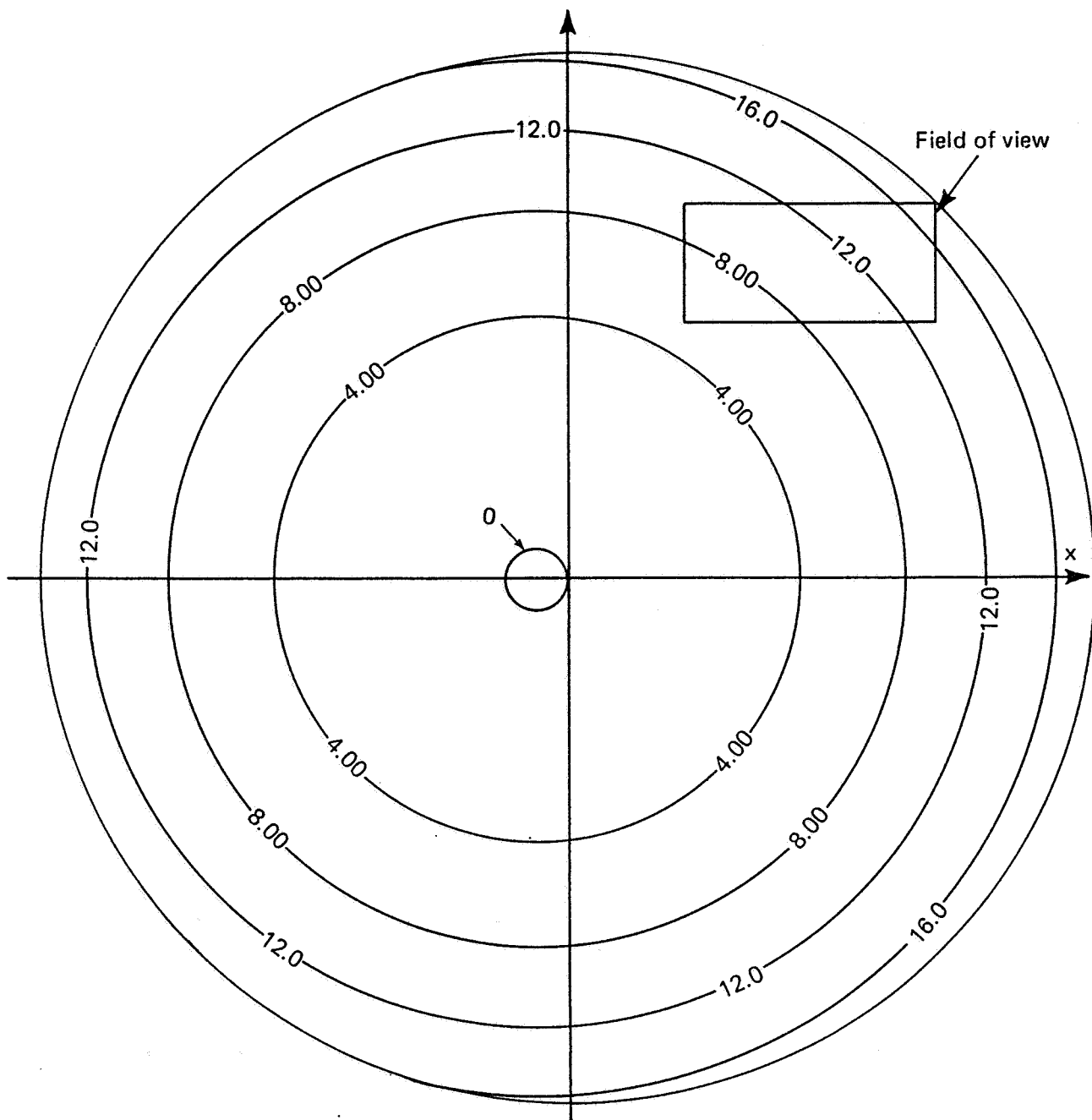


Fig. C.6. Isovelocities due to the combined effect of the Sun's rotation and the untilted FP. The rectangle designates the field of view. Isovelocities in km/s^{-1} .

In our case $\Delta X_O = 1/2$, $\Delta Y_O = 1/4$.

$$\Delta V = v_E/2 + 2d(X_O/2 + Y_O/4) . \quad (25)$$

Using some algebra, we can put an upper limit on the sum $(X_O/2 + Y_O/4)$

$$(X_O/2 + Y_O/4) < 0.4 . \quad (26)$$

So the upper limit on ΔV is

$$\Delta V < \frac{v_E}{2} + 0.8\alpha = 13.28 \text{ km s}^{-1} . \quad (27)$$

The ratio of this maximum variation to the width of most solar lines is about one. This is unacceptable, so we will minimize the velocity variation by tilting the FP using the technique described by Rust et al. (1986b). It is easy to show that the minimum velocity variation is given by:

$$\Delta V_{\min} = \alpha \left(\left(\frac{\Delta X_O}{2} \right)^2 + \left(\frac{\Delta Y_O}{2} \right)^2 \right) . \quad (28)$$

With the values given above $\Delta V_{\min} = 1.2 \text{ km s}^{-1}$. To obtain this value the FP will be tilted in two directions X and Y. The minimalization could be done in real time if the observer is provided with an appropriate display.

3. Performance in an Imaging System

The optics of the magnetograph must provide collimated light to the FP filter at very small angles of incidence while maintaining a useful field of view and high spatial resolution. Figure C.7 is a schematic drawing of an optical train that we believe can achieve these objectives and the low instrumental polarization required for vector field measurements. The telescope is an existing f/6.11 Ritchey-Chretien with a 33-cm-diameter primary mirror. The effective aperture of the telescope is 31.2 cm. The central obstruction of the reflecting system is 50% in diameter. At a wavelength of 5250 Å, the telescope resolution is limited by diffraction to 0.42 arc sec over the ± 4.3 arc min field of view, which will be defined by a stop at the telescope focus. The etendue (effective aperture x field of view) is 0.1994 m²-ster. This figure takes into account a reduction in the field of view to 4.3 x 8.5 arc min at the detector.

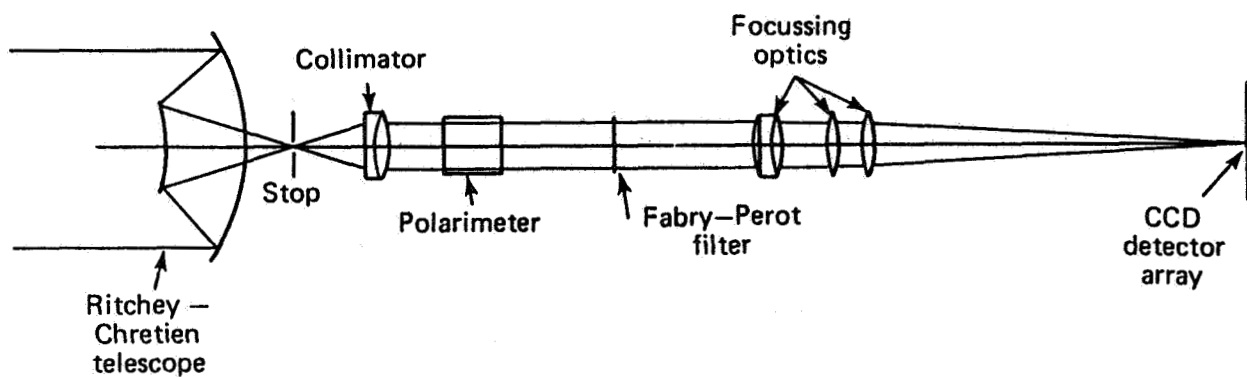


Fig. C.7. Magnetograph optical system layout.

Beyond the stop at the primary focus is a telecentric system consisting of a refractive doublet that projects collimated light through the polarization analyzer and the FP filter. An image of the pupil is located at the FP aperture. The collimator provides light to the FP at an angle of incidence range of ± 6.3 mrad. Neither the collimator nor the FP restricts the 0.195 mm-rad Lagrangian invariant (aperture radius \times field of view half angle) of the system. Behind the FP a final series of refractive elements focuses the image on the detector array. The overall resolution is very near the diffraction limit within the defined field of view. The overall length of the system is 3.5 m and the effective focal length at the detector is 23.2 m.

4. Survival in Space

An instrument in space is insulted by a variety of radiations and operational stresses that may limit its useful life. First is the vibration of launch, which conceivably could shatter the etalon, since the thickness of the lithium niobate substrate, which is more brittle than glass, is only 0.2 mm. Therefore, we intend to expose sample wafers of lithium niobate to simulated launch stresses. No difficulty is anticipated in space-qualifying an etalon, however, because standard techniques exist for protecting thin, brittle membranes during launch. The basic approach is to surround the membrane with acoustic-wave absorbing material.

The potential for radiation damage in lithium niobate depends on the orbit of the spacecraft. For equatorial orbits, one encounters energetic electrons, principally, and the etalon can be shielded from them. For a spacecraft outside the protecting envelope of the Earth's magnetic field, energetic protons could be a problem, since there is no practical shield against them. By disrupting the crystalline structure of the niobate, energetic protons could cause optical damage and lower the transmission of the filter. However, experiments on lithium niobate samples (Allen et al., 1972) showed no change in transmission or index of refraction after exposure to 2×10^5 rads (AL) (1.8×10^{12} protons cm^{-2} at $E = 85$ MeV or 3.4×10^{12} protons cm^{-2} at $E = 208$ MeV), which is an order of magnitude more radiation than expected on a 5-year mission at L_1 with a minimum 1-mm-thick aluminum shield (JHU/APL, 1985). Lithium niobate was also exposed to 6.4×10^7 rads (AL) ($\sim 3 \times 10^{15}$ e cm^{-2}) in a beam of 2.2-MeV electrons with no change in transmission or refractive index. In these tests, the optical properties were measured at 5300 Å.

Damage to lithium niobate caused by optical beams has been studied extensively (Glass et al., 1972) and could be a problem in an etalon exposed to direct sunlight for an extended period. In a spaceborne magnetograph, prefilters will remove the damaging UV and most of the unwanted light. Solar optical radiation at ~ 5000 Å amounts to $0.02 \text{ mW cm}^{-2} \text{ Å}^{-1}$, so with a telescope collecting area of 640 cm^2 , a passband of 100 Å, and transmission

efficiency of ~25% up to the FP (whose area is 30 cm^2), the optical flux on the FP will be 10 mW cm^{-2} . Since the first mirror surface will reflect 90% of this, the flux inside the etalon would be only 1 mW cm^{-2} were it not for the fact that the internal reflections amplify the beam there by a factor of F. With the etalon described above, $F = 25$, so the internal optical flux is 25 mW cm^{-2} . This is still much lower than the thresholds (300 W cm^{-2}) at which optical damage has been observed with laser beam illumination.

In order to avoid distortion of the optics and to subject the etalon to a minimum thermal load, infrared radiation will be rejected by a prefilter over the face of the telescope. Most of the 30 mW of solar light passing through the first mirror surface is either transmitted by the etalon or reflected. Very little energy is absorbed, even by the silver coatings. Absorption can be further minimized by using thin film dielectrics for the reflective coatings. These films absorb less energy than metallic films and their long-term durability has been established (Title et al., 1974).

In order to make the magnetograph as durable as possible, mechanical devices should be avoided, including image scanners, rotating filter elements, and moving polarization analyzers. A two-dimensional, solid-state charge-coupled device (CCD) is the detector of choice since it can record the entire 4.3×8.6 arc min field of view without mechanical scanning. All the magnetic flux in even the largest activity centers that appear on the disk can be recorded at 0.25 arc sec per pixel with a 2048×2048 CCD. CCD cameras have been space qualified in a number of programs, e.g., Galileo.

5. Final Remarks

As the signal-to-noise analysis in section C.II shows, long exposure times are unavoidable if magnetic field sensitivity of a few tens of gauss is to be achieved at half-arc sec spatial resolution with a moderate size telescope. The now-cancelled 1.25-m Solar Optical Telescope still represents the ultimate solution to the severe requirements of high-resolution solar magnetic field detection, but we believe that substantial progress in solar research can be achieved by operating the vector magnetograph described here at a ground site that offers at least occasional periods of 0.5 to 1 arc sec images. Another approach is to operate the instrument on a long-duration balloon flight at an altitude of $\sim 30 \text{ km}$, where atmospheric blurring is no longer a factor.

To achieve its full promise, the FP filter should be provided with a tunable blocking filter, so that different orders can be selected quickly without having to rotate a filter wheel. One possibility for a tunable blocker is the acousto-optic filter (Harris and Wallace, 1969). Electronic tuning of the device is accomplished by tuning an acoustic wave generator.

Another possibility for a tunable blocker is the liquid-crystal filter described by Gunning et al. (1981). We have already measured the properties of a prototype commercial liquid-crystal device (LCD) (Rust, 1985). A modification of the traditional LCD design is needed to make a tunable blocker. A 25- μm layer of liquid crystal sandwiched between FP mirrors should produce a 1- \AA filter that can be tuned by application of a few volts to the mirror faces. The device is optically similar to the lithium niobate FP; its free spectral range is much higher, however, making it an efficient blocker for the lower FSR device.

6. Addendum - Vector Field Measurement Uncertainties

The purpose of this Addendum is to determine the errors in measurements of the vector magnetic field. The simple case we will evoke is the weak field case.

The Stokes parameters are defined as follows:

$$I = I_{0^\circ} + I_{90^\circ} = I_+ + I_- \quad (\text{A1})$$

$$Q = I_{0^\circ} - I_{90^\circ} \quad (\text{A2})$$

$$U = I_{-45^\circ} - I_{+45^\circ} \quad (\text{A3})$$

$$V = I_+ - I_- \quad (\text{A4})$$

where I_{-45° , I_{0° , and I_{90° are the intensities with the indicated phase retardation. I_+ and I_- are the circular polarized intensities. I , Q , U , and V are functions of the position (or wavelength offset) of the filter (here the FP) compared to the center of the line. They depend also on the line profile, the magnetic field, and the angle that the magnetic field makes with the line of sight (e.g., Richter et al., 1985).

$$I(v_O) = 1/4 (1 + \cos^2 \gamma) [p(v_O + v_B) + p(v_O - v_B)] + 1/2 \sin^2 \gamma p(v_O) \quad (\text{A5})$$

$$Q(v_O) = 1/4 \sin^2 \gamma \cos 2\alpha [2p(v_O) - p(v_O + v_B) - p(v_O - v_B)] \quad (\text{A6})$$

$$U(v_O) = 1/4 \sin^2 \alpha \sin 2\alpha [2p(v_O) - p(v_O + v_B) - p(v_O - v_B)] \quad (\text{A7})$$

$$V(v_O) = 1/2 \cos \gamma [p(v_O + v_B) - p(v_O - v_B)] , \quad (\text{A8})$$

where v_0 is the filter passband offset relative to the width Δv_0 of the line, $p(v)$ is the convolved profile of the line and the filter (FP), v_B is the equivalent magnetic shift relative to Δv_0 ($v_B \Delta v_0 = g B/1.2$, where g is the Lande factor), γ is the inclination of the magnetic field vector, and α is the azimuth angle.

In the weak field case ($v_B \ll 1$) we can approximate these parameters to second order in v_B by:

$$I(v_0) = p(v_0) + O(v_B) \quad (A9)$$

$$Q(v_0) = 1/4 \sin^2 \gamma \cos 2\alpha v_B^2 \left. \frac{\partial^2 p}{\partial v^2} \right|_{v=v_0} + O(v_B^4) \quad (A10)$$

$$U(v_0) = -1/4 \sin^2 \gamma \sin 2\alpha v_B^2 \left. \frac{\partial^2 p}{\partial v^2} \right|_{v=v_0} + O(v_B^4) \quad (A11)$$

$$V(v_0) = \cos \gamma v_B \left. \frac{\partial p}{\partial v} \right|_{v=v_0} + O(v_B^3) \quad (A12)$$

Define the reduced Stokes parameters as:

$$q(v_0) = 1/4 \sin^2 \gamma \cos 2\alpha \left. \frac{\partial^2 p}{\partial v^2} \right|_{v=v_0} \frac{1}{p(v_0)} v_B^2 \quad (A13)$$

$$u(v_0) = -1/4 \sin^2 \gamma \sin 2\alpha \left. \frac{\partial^2 p}{\partial v^2} \right|_{v=v_0} \frac{1}{p(v_0)} v_B^2 \quad (A14)$$

$$v(v_0) = \cos \gamma v_B \left. \frac{\partial p}{\partial v} \right|_{v=v_0} \frac{1}{p(v_0)} \quad (A15)$$

We can solve these equations to obtain γ , α , v_B or $v_{||}$, v_{\perp} , and α .

Assume

$$P_1 = \frac{1}{p(v_0)} \left. \frac{\partial p}{\partial v} \right|_{v=v_0}, \quad P_2 = \frac{1}{p(v_0)} \frac{1}{4} \left. \frac{\partial^2 p}{\partial v^2} \right|_{v=v_0} \quad (A16)$$

$$v_{\parallel} = v(v_o)/P_1$$

$$v_{\perp} = \frac{1}{P_2^{1/2}} [u^2(v_o) + q^2(v_o)]^{1/4} \quad (A17)$$

$$\tan 2\alpha = u/q$$

6.1 Uncertainties

A very crude way to calculate the rms noise or uncertainties in these measurements is to assume that Δu , Δv , and Δq , are given by the photon noise.

$$\Delta u = \Delta v = \Delta q = \frac{1}{\sqrt{N}} \frac{1}{\sqrt{p(v_o)}} \quad (A18)$$

where N is the number of photons collected by a rectangular window of width Δv_o in unpolarized light and $p(v_o)$ is the convolved profile $\{I(y/2)$ in the main text).

$$\Delta v_{\parallel}^{\text{rms}} = \frac{1}{\sqrt{N}} \frac{\sqrt{p(v_o)}}{\left. \frac{\partial p}{\partial v} \right|_{v=v_o}} \frac{1.2}{g} \frac{\Delta v_o}{\sqrt{N}} \sqrt{2} Q(u, y). \quad (A19)$$

This expression is about the same as given for δv_B^{rms} except that we make only one measurement, at $v = v_o$ (i.e., at $y/2$):

$$B_{\parallel}^{\text{rms}} = \frac{1.2}{g} \frac{\Delta v_o}{\sqrt{N}} \frac{\sqrt{p(v_o)}}{\left. \frac{\partial p}{\partial v} \right|_{v=v_o}}. \quad (A20)$$

$$\Delta v_{\perp}^2 = \frac{1}{P_2^{1/2}} \left[\frac{2|u(v_o)|\Delta u + 2|q(v_o)|\Delta q}{2[u(v_o) + q^2(v_o)]^{1/2}} \right] \quad (A21)$$

or

$$u(v_o) = v_{\perp}^2 P_2 \sin 2\alpha \quad (A22)$$

$$q(v_o) = v_{\perp}^2 P_2 \cos 2\alpha \quad (A23)$$

$$\Delta v_{\perp}^2 = \frac{1}{P_2^{1/2}} \frac{(|\sin 2\alpha| + |\cos 2\alpha|)}{\sqrt{N} \sqrt{P(v_o)}} \quad (A24)$$

$$\Delta v_{\perp} = \frac{1}{v_{\perp}} \frac{(|\sin 2\alpha| + |\cos 2\alpha|)}{\sqrt{N}} \left[\frac{\partial^2 p}{\partial v^2} \right]_{v=v_o}^{-1/2} \quad (A25)$$

For weak fields, we may have $\Delta v_{\perp} > v_{\perp}$ unless a great many measurements are averaged. Let us estimate Δv_{\perp} when $\Delta v_{\perp} = v_{\perp}$, the upper limit on Δv_{\perp} is given by:

$$\Delta v_{\perp} \leq \Delta v_{\perp \text{MAX}} = \left(\frac{2}{N}\right)^{1/4} \left[\frac{\partial^2 p}{\partial v^2} \right]_{v=v_o}^{-1/4} \quad (A26)$$

Using (9) we find that B_{\perp}^{rms} , the upper limit on the rms detectable transverse magnetic field, is:

$$B_{\perp}^{\text{rms}} = \frac{1.2}{g} \Delta v_o \left(\frac{2}{N}\right)^{1/4} \left[\frac{\partial^2 p}{\partial v^2} \right]_{v=v_o}^{-1/4} \quad (A27)$$

It means that we cannot measure transverse magnetic fields weaker than B_{\perp}^{rms} with a relative precision better than 100% a factor of two. It must be noted that B_{\perp}^{rms} is decreasing much more quickly than $B_{\parallel}^{\text{rms}}$, when N is increasing. In our case B_{\perp}^{rms} will be about 300 gauss for a 1-sec integration time (see Figure C.8).

$$\Delta \tan 2\alpha = \frac{2\Delta\alpha}{\cos^2 2\alpha} = \frac{\Delta u}{q} + u \frac{\Delta q}{q^2} \quad (A28)$$

$$\Delta\alpha = \frac{\cos^2 2\alpha}{2} \Delta u \left(\left| \frac{1}{q} \right| + \left| \frac{u}{q^2} \right| \right)$$

$$\begin{aligned} &= \frac{\cos^2 2\alpha}{2} \Delta u \left(\frac{1}{v_{\perp}^2 P_2 |\cos 2\alpha|} + \frac{|\tan 2\alpha|}{v_{\perp}^2 P_2 |\cos 2\alpha|} \right) \\ &= \frac{\cos^{1/2} 2\alpha}{2} \Delta u \frac{|\sin 2\alpha| + |\cos 2\alpha|}{v_{\perp}^2 P_2} \end{aligned} \quad (A29)$$

$$\Delta\alpha = \frac{1}{2} \frac{1}{\sqrt{N}} \frac{1}{\sqrt{P(v_o)}} \frac{1}{v_{\perp}^2} [|\sin 2\alpha| + |\cos 2\alpha|] 4P(v_o) \left[\frac{\partial^2 p}{\partial v^2} \right]_{v=v_o}^{-1}$$

$$\Delta\alpha = \frac{2}{\sqrt{N}} \frac{\sqrt{p(v_o)}}{v_{B1}^2} \left[\frac{\partial^2 p}{\partial v^2} \right]_{v=v_o}^{-1} (|\sin 2\alpha| + |\cos 2\alpha|) . \quad (A30)$$

The upper limit on $\Delta\alpha$ is given by:

$$\Delta\alpha < \Delta\alpha_{MAX} = \frac{2\sqrt{2}}{\sqrt{N}} \frac{\sqrt{p(v_o)}}{v_{B1}^2} \left[\frac{\partial^2 p}{\partial v^2} \right]_{v=v_o}^{-1} . \quad (A31)$$

$\Delta\alpha_{MAX}$ could be approximated by:

$$\Delta\alpha_{MAX} \approx \frac{2\sqrt{2}}{\sqrt{N}} \frac{1}{v_{B1}^2} . \quad (A32)$$

In our case with v_{B1} equivalent to 300 gauss and 100-sec integration time we get $\Delta\alpha_{MAX} \approx 10^0$. Figure C.8 shows that precise measurements of weak field are very difficult.

Vector field measurements are easier when the fields are not weak. This is because U and Q are proportional to B^2 , while V is proportional to B ; for a strong field, the U , Q , and V signals become comparable.

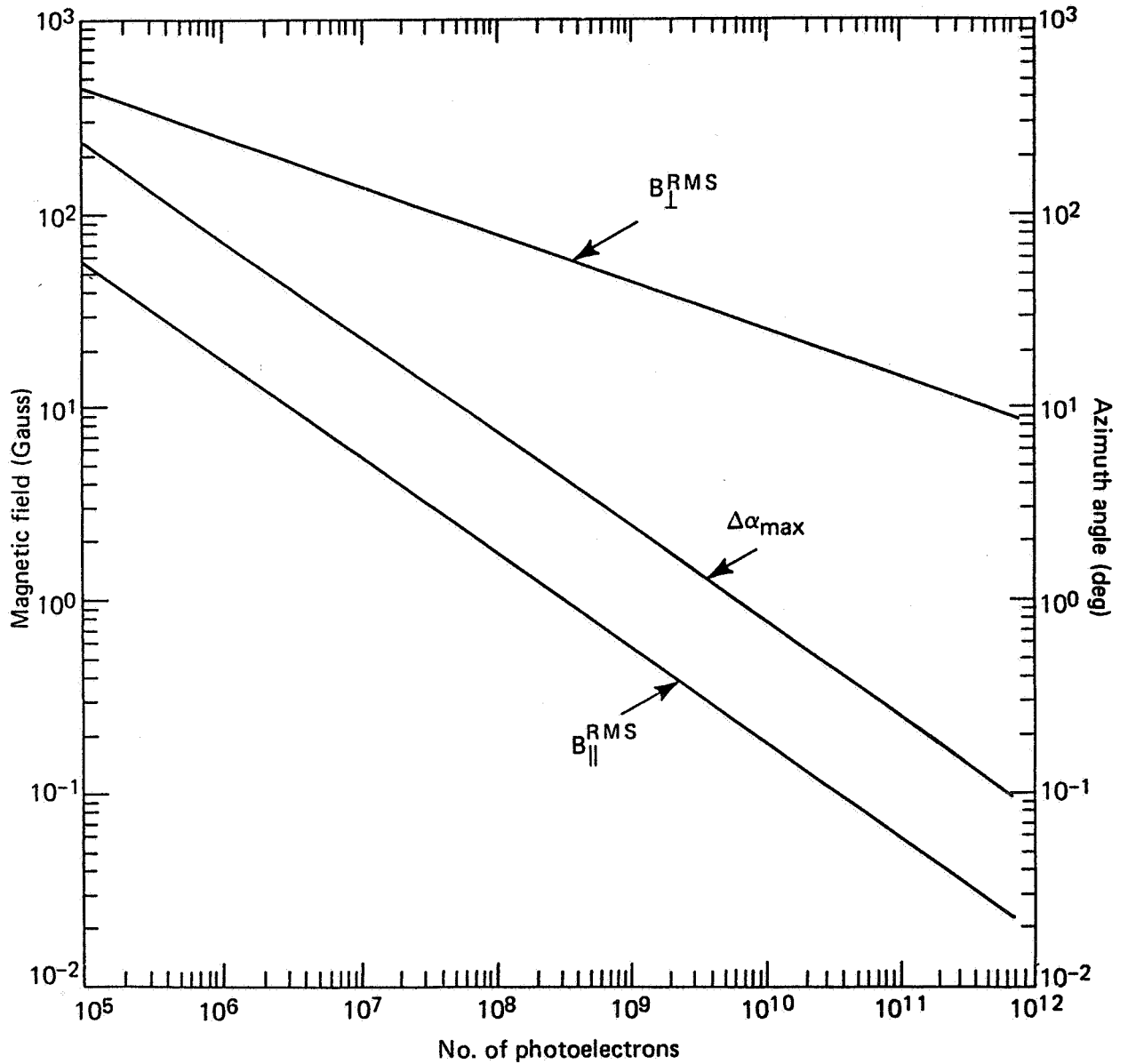


Fig. C.8. Errors on the measurements of B_{\parallel} , B_{\perp} and α . These curves were computed assuming a line width $\Delta\nu_0 = 0.15\text{\AA}$, a depth $P = 0.5$, and a Landé factor $g = 1.5$. The quality factor Q was assumed to be about 2. $\partial^2 p / \partial \nu^2|_{\nu=0}$ was assumed to be about 1. The magnetic field taken to compute $\Delta\alpha_{max}$ is $B_{\perp} = 300$ gauss.

ACKNOWLEDGMENTS

This work was supported by the NASA Marshall Space Flight Center, the APL Independent Research and Development Program, and by NASA Grant NAGW-522. We are grateful to Dr. Clive Burton of the Australian Commonwealth Scientific and Industrial Research Organization for supplying information about the etalons and to Professor J. O. Stenflo of ETH in Zurich for helpful discussions.

REFERENCES

- L. B. Allen et al., "Electro-Optical Modulator," DTIC Tech. Rept. AFAL-TR-72-60, March 1972.
- A. M. Glass, G. E. Peterson, and T. J. Negran, "Optical Index Damage in Electro-Optical Crystals," in Laser Induced Damage in Optical Materials NBS Pub. 372 (1972) Washington, D.C. pp. 15-26.
- W. Gunning, "Double-Cavity Electrooptic Fabry-Perot Tunable Filter," Appl. Optics, 21, 3129 (1982).
- W. Gunning, J. Pasko, and J. Tracy, "A Liquid Crystal Tunable Spectral Filter: Visible and Infrared Operation," in SPIE Proc. 268, p. 190 (1981).
- M. J. Hagyard (ed.), Measurements of Solar Vector Magnetic Fields, NASA Conf. Pub. 2374, p. 109 (1985).
- S. E. Harris and R. W. Wallace, "Acousto-Optic Tunable Filter," J. Opt. Soc. Amer., 59, 744 (1969).
- JHU/APL, "Conceptual Design of a Synoptic Interplanetary Monitor Plat format L1 (SIMPLE)," in The Johns Hopkins University Applied Physics Laboratory, Doc. SDO 7805, November 1985, p. 156.
- S. F. Martin, S. H. B. Livi, and J. Wang, "The Cancellation of Magnetic Flux. II. In a Decaying Active Region," Australian J. Physics, 38, 929 (1985).
- P. H. Richter, L. K. Zeldin, and T. A. Loftin, "The San Fernando Observatory Video Stokes Polarimeter," in Measurements of Solar Vector Magnetic Fields, NASA Conf. Publ. 2374 (1985).
- D. M. Rust, "Some Design Considerations for a Satellite-Borne Magnetograph," in Measurements of Solar Vector Magnetic Fields, NASA Conf. Publ. 2374 (1985).
- D. M. Rust and C. A. Bridges, "The Work of the Diode Array: He 10830 Observations of Spicules and Subflares," Solar Physics 43, 129 (1975).
- D. M. Rust, C. H. Burton, and A. J. Leistner, "A Tunable Solid Fabry-Perot Etalon for Solar Seismology," in SPIE Proc. 627-Instrumentation in Astronomy VI, p. 39 (1986a).
- D. M. Rust, T. Appourchaux, and F. Hill, "Performance of a Stabilized Fabry-Perot Solar Analyzer" in Advances in Helio- and Astroseismology, IAU Symp. 123, ed. J. Christensen-Dalsgaard, Dordrecht: Reidel (1986b).

- D. M. Rust, R. F. Cohn, and R. Kunski, "Development of Ultra-Stable Filters and Lasers for Solar Seismology," APL Technical Digest 7, 209 (1986c).
- A. M. Title, T. P. Pope, and J. P. Andelin, Jr., "Drift in Interference Filters," Appl. Optics, 13, 2675 (1974).

APPENDIX D

Polarization Aberration Theory

1. Introduction

This Appendix describes the method which has been developed to calculate the instrumental polarization of the SAMEX optics. A full development of the method is contained in "Polarization Aberrations" by Chipman (1987). The calculation requires performing thin film calculations during the optical design process to determine the two parts (eigenvectors) of an optical beam, as a function of the object and pupil coordinates.

The mathematical method of treating instrumental polarization will start with a discussion of the Jones and C vectors and conclude with a discussion of the polarization aberration expansion and a derivation of the polarization accuracy.

2. The Jones Matrix and C Vector for the Characterization of Polarization

The most efficient mathematical method for treating the SAMEX instrumental polarization is the Jones calculus. The Mueller calculus is a more difficult representation which includes optical depolarization (scatters) properties. Such weak depolarization effects are more readily handled by experimental measurements. The Jones calculus (Jones, 1941a,b,c, 1942, 1947; Azzam and Bashara 1977; Theocaris and Gdoutos, 1979) is a mathematical formalism for treating problems involving the description of polarized light and polarizers which uses the Jones vector for the description of polarized light and the Jones matrix to characterize the polarizing properties of an optical element. The details of the following discussion are given by Chipman (1987). The elements of the Jones matrix and the C vector characterization are outlined first.

2.1 Definition of the Jones Vector in terms of the Electric Field Amplitudes

The Jones vector expression for a quasi-monochromatic plane wave propagating parallel to the z axis with electric field amplitude $\vec{E}(t)$ is

$$\vec{E}(t) = \vec{E}_x(t) + \vec{E}_y(t) \quad ,$$

where

$$\vec{E}_x(t) = \hat{n}_x E_{0,x}(t) \cos[(kz - \omega t) + e_x] \quad ,$$

and,

$$\vec{E}_y(t) = \hat{n}_y E_{0,y}(t) \cos[(kz - \omega t) + e_y] ,$$

where \hat{n}_x and \hat{n}_y are unit vectors in the x and y direction with the light propagation in the z direction. There are four parameters, $E_{0,x}$, $E_{0,y}$, e_x , and e_y , defining the wave besides the wavelength.

The time-dependent Jones vector is defined in terms of the electric field amplitudes as

$$\vec{J}(t) = \begin{bmatrix} E_x(t) \\ E_y(t) \end{bmatrix} .$$

The components of $\vec{J}(t)$ are the instantaneous components of $\vec{E}(t)$.

The normalized Jones vector \vec{J} is a time-independent normalized vector where all the vector components of $\vec{J}(t)$ have been divided by the incident electric field amplitude,

$$\vec{J} = \vec{J}(t)/E_0(t) .$$

The normalized Jones vector is referred to as "the Jones vector" unless otherwise stated. Knowledge of \vec{J} and E_0 provides all the information necessary to reconstruct $\vec{E}(t)$ to within a constant phase factor.

Table D.1 lists the Jones vectors for the most common polarization states: horizontal linear, vertical linear, $+45^\circ$ linear, -45° linear, right circular and left circular polarized light. These vectors can be multiplied by an arbitrary phase factor without changing the polarization ellipse of the light; it only changes the absolute phase.

2.2 The Jones Matrix

Having established the vector which defines the polarization state we now consider the matrices which represent the polarization effect of the optical elements and allow polarization calculations to be performed.

Table D.1. The basic Jones vector representation for linear and circular polarized light.

<u>Linear Polarized Light</u>			
Horizontal $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$	Vertical $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$	+45° s $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$	-45° s $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$

<u>Circular Polarized Light</u>	
Right Circular s $\begin{pmatrix} 1 \\ i \end{pmatrix}$	Left Circular s $\begin{pmatrix} 1 \\ -i \end{pmatrix}$

where $s = \frac{\sqrt{2}}{2}$

In his original paper, Jones (1941a) showed that the relationship between the Jones vector incident on a polarizer, \vec{J} , and that transmitted or reflected by a polarization element, \vec{J}' , can always be related by a matrix, the Jones matrix, JJ . Only certain transformations of the field components are allowed, those described by a matrix. Thus the fundamental relationship between the vector components of the electromagnetic fields before and after a polarizing element is,

$$\vec{J}' = JJ \cdot \vec{J} .$$

The Jones matrix, JJ , is a 2 by 2 matrix with complex elements,

$$JJ = \begin{bmatrix} j(1,1) & j(1,2) \\ j(2,1) & j(2,2) \end{bmatrix} ,$$

where $j(k,1) = a(k,1) + i b(k,1) .$

Thus the Jones matrix has eight degrees of freedom. Thus there are eight different forms of polarization behavior. These eight forms are listed in Table D.2. Every Jones matrix corresponds to a physically realizable polarizer.

2.2.1 Sequences of polarizers

The Jones matrix associated with an optical ray path through a sequence of polarization elements is just the matrix product of the Jones matrices for the individual polarizers. If an optical ray traverses a series of elements, 1, 2, ... Q, and the Jones matrices appropriate to that ray for each element are JJ(1), JJ(2), ... JJ(Q), then the Jones matrix describing the polarization properties of the system along this ray path is given by the matrix product,

$$JJ = JJ(Q) \dots JJ(2) JJ(1)$$

Since the Jones matrix of an optical element is dependent upon the wavelength, angle of incidence, orientation, and path through the element, each ray in each wavelength will usually have a different Jones matrix. Only if a collimated monochromatic beam through a series of planar optical interfaces can be assumed, then a single Jones matrix can be written for the entire cross section of the beam. This is the case in the polarimeter section of the magnetograph but not in the foreoptics.

2.2.2 Coordinate system

The Jones matrix is defined relative to an arbitrary x and y coordinate system. Since these coordinates have been defined for the Jones vector, the coordinate system of the Jones matrix is defined in terms of the Jones vector coordinates.

It is often desirable to align the Jones vectors coordinates with the s and p planes of an optical interface. Only for plane surfaces does the orientation of the s and p planes remain fixed across the surface. For non-planar surfaces, it is necessary to maintain two sets of coordinates, the global x and y coordinates with respect to which the Jones matrix is defined, and a local s and p coordinate for each individual point on the interface. The local s and p coordinate system will have its x' and y' axes aligned with the local s and p planes of the surface for the evaluation of Jones matrices at given surface coordinates. Then, these local Jones matrices will be rotated to bring all the matrices into the global coordinate system.

2.2.3 Pauli spin matrix basis and the C vector

The Pauli spin matrices form a most useful basis for interpreting the Jones matrix "space" and define a basis set for the JJ matrix. The identity matrix, $\sigma(0)$, and the Pauli spin matrices, $\sigma(1)$, $\sigma(2)$, and $\sigma(3)$, are defined by:

$$\begin{pmatrix} \sigma(0) \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \begin{pmatrix} \sigma(1) \\ 1 & 0 \\ 0 & -1 \end{pmatrix} \quad \begin{pmatrix} \sigma(2) \\ 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \begin{pmatrix} \sigma(3) \\ 0 & -\tilde{i} \\ \tilde{i} & 0 \end{pmatrix} .$$

An arbitrary Jones matrix will be expressed as

$$JJ = \sum_{k=0}^3 c(k) \sigma(k).$$

The c's are formed into a four-element complex vector, the "C vector." The C vector expression

$$C = [c(0), c(1), c(2), c(3)]$$

represents the Jones matrix

$$JJ = c(0) \sigma(0) + c(1) \sigma(1) + c(2) \sigma(2) + c(3) \sigma(3)$$

$$= \begin{pmatrix} c(0) + c(1) & c(2) - \tilde{i}c(3) \\ c(2) + \tilde{i}c(3) & c(0) - c(1) \end{pmatrix} .$$

When necessary, ρ and ϕ refer to the amplitude and phase portions of the C vector elements:

$$C = [\rho(0) \exp(\tilde{i}\phi(0)), \rho(1) \exp(\tilde{i}\phi(1)), \rho(2) \exp(\tilde{i}\phi(2)), \rho(3) \exp(\tilde{i}\phi(3))].$$

The elements of C are related to the Jones matrix elements by the equations:

$$\begin{aligned} c(0) &= (j(1,1) + j(2,2))/2, & c(1) &= (j(1,1) - j(2,2))/2, \\ c(2) &= (j(1,2) + j(2,1))/2, & c(3) &= (j(1,2) - j(2,1))/(-2\tilde{i}). \end{aligned}$$

The elements of the Jones matrix are related to the elements of C by the equations:

$$\begin{aligned} j(1,1) &= c(0) + c(1), & j(1,2) &= c(2) - \tilde{i}c(3), \\ j(2,1) &= c(2) + \tilde{i}c(3), & j(2,2) &= c(0) - c(1). \end{aligned}$$

The C vector, like the Jones matrix, has eight degrees of freedom. Table D.2 contains a description of the meaning of the real and imaginary parts of the C vector elements.

TABLE D.2. Interpretation of the C vector elements.

Matrix	Coefficient	Meaning	
$\sigma(0)$	$\rho(0)$	Amplitude	Absorption
$\sigma(0)$	$\phi(0)$	Phase	Phase
$\sigma(1)$	$\rho(1)$	Amplitude	Linear Polarization along Axes
$\sigma(1)$	$\phi(1)$	Phase	Linear Retardance along Axes
$\sigma(2)$	$\rho(2)$	Amplitude	Linear Polarization, 45°
$\sigma(2)$	$\phi(2)$	Phase	Linear Retardance, 45°
$\sigma(3)$	$\rho(3)$	Amplitude	Circular Polarization
$\sigma(3)$	$\phi(3)$	Phase	Circular Retardance

2.2.4 The Jones matrix and C vectors for specific polarizers

Tables of Jones matrices for various polarizers are found in Azzam and Bashara (1977, Section 2.2.3), Hecht and Zajac (1974, Table 8.6), Shurcliff (1962, Appendix 2), and Theocaris and Gdoutos (1979, Table 4.1). A table listing of the Jones matrices and C vectors for the most common polarizers and retarders is given in Chipman (1987, Table 6).

2.2.5 The meaning of the coefficients of the C vector

The primary reason for the introduction of the C vector is to simplify the representation of polarizers. Each of the elements of the C vector represents a specific type of polarizer behavior.

The real parts of the C vector all correspond to amplitude effects, absorption, and dichroism. The phase portion of the C vector represents phase effects, propagation, and birefringence. The first element, $c(0) = \rho(0)\exp(i\phi(0))$, is the coefficient of the identity matrix. Thus it must represent effects that are independent of polarization state; these are amplitude and phase. The last element, $c(3) = \rho(3)\exp(i\phi(3))$, multiplies the spin matrix $\sigma(3)$ which is rotation invariant. Thus the $c(3)$ term represents the circular polarization

effects, $\rho(3)$ describes circular polarization or circular dichroism, and $\phi(3)$ describes circular retardance or circular birefringence. The remaining two elements, $c(1)$ and $c(2)$, represent linear polarization and linear retardance. Linear terms require two degrees of freedom: magnitude and orientation. Thus, $\rho(1)$ and $\rho(2)$ characterize linear polarization or linear dichroism, $\rho(1)$ in the 0° and 90° directions, $\rho(2)$ in the $\pm 45^\circ$ directions. Likewise, $\phi(1)$ and $\phi(2)$ characterize linear retardance or linear birefringence.

2.2.6 Rotated polarizers

If a polarizer with Jones matrix JJ is rotated through an angle θ (positive if counterclockwise), the Jones matrix becomes

$$JJ'(\theta) = R(\theta) JJ R(-\theta).$$

The $R(\theta)$'s are the Jones rotation matrices:

$$R(\theta) = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}.$$

The Jones rotation matrices obey the relations

$$R(a) R(b) = R(b) R(a) = R(a+b)$$

and

$$R(a) R(-a) = \sigma(0).$$

The identity matrix is invariant under rotation:

$$R(\theta) \sigma(0) R(-\theta) = \sigma(0).$$

Under rotation, $\sigma(1)$ and $\sigma(2)$ couple into each other:

$$R(\theta) \sigma(1) R(-\theta) = \sigma(1)\cos(2\theta) + \sigma(2)\sin(2\theta)$$

$$R(\theta) \sigma(2) R(-\theta) = -\sigma(1)\sin(2\theta) + \sigma(2)\cos(2\theta).$$

$\sigma(3)$ is invariant under rotation: $R(\theta) \sigma(3) R(-\theta) = \sigma(3).$

Having established the polarization calculus which describes the polarization optics, we now apply these matrices to the SAMEX optical system.

3. Instrumental Polarization

Two types of polarization calculations can be performed for the SAMEX magnetograph: instrumental polarization and transmitted light polarization. The first is the calculation of the polarization associated with ray paths through an optical system. This determines the instrumental polarization function as a function of pupil coordinates for a specified object point. The other type of calculation determines the state of polarization, such as a Jones vector, transmitted by the system along a given ray path for a specified input polarization state. By iterating this process, the Jones vector as a function of position in the exit pupil is calculated. This Appendix deals only with the instrumental polarization calculation. Once the instrumental polarization function for the system is known, the transmitted Jones vectors are readily determined for all input polarization states.

Polarizers are optical elements which divide an optical beam into two parts (Jones vector) and transmit those parts with a different transmission coefficient and a different phase. The two parts of the beam are referred to as the eigenvectors or as the more descriptive term, "eigenpolarizations." The two eigenpolarizations are orthogonally polarized and are transmitted by the polarizer with no alteration of their polarization states; only the intensity and phase change.

The word polarizer will be used to refer to both polarizers, such as dichroic or prism types which have a different transmittance for the two eigenpolarizations, and retarders, which have equal transmittance but a different phase change for the polarizations. Polarized Light by Shurcliff(1962) is the standard reference on the types of polarizers, their definitions, parameters, and properties.

3.1 Transparent Systems

The SAMEX foreoptics are highly transparent and weakly polarizing and the following calculational method is optimized for this case. The ideal Jones matrix for a ray through a transparent nonpolarizing system is

$$JJ(\text{ideal}) = \exp(i\tilde{d}) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

where d is the optical path length for the ray in radians. (The complex value $(-1)^{1/2}$ is denoted by i .) The Jones matrix operation on the Jones vector, which is composed of the two orthogonal amplitude components, defines the transmitted state of the ray.

The above ideal Jones matrix is the identity matrix, which signifies that the system has no absorption or polarization. Since this is the desired form of the Jones matrix for the SAMEX foreoptics, the approach developed here obtains the instrumental polarization function as a Taylor series of the system Jones matrix in the ray coordinates about JJ(ideal). This approach is easily modified for systems which are not highly transparent or which contain strong polarizers by performing the Taylor series about the Jones matrix for the ray down the optical axis. This work deals primarily with this simpler version of the problem, transparent systems, to streamline the SAMEX calculations.

3.2 S-P Coordinates

To handle problems involving light at non-normal incidence at curved optical surfaces, it is necessary to maintain two separate coordinate systems: x-y coordinates and s-p coordinates. The x-y coordinates are the global x, y, and z coordinate system used to describe the optical system. The optical axis of radially symmetric optical systems coincides with the z axis.

The s-p coordinates are used to perform polarization calculations with the SAMEX coatings which depend on the angle of incidence. Most frequently, the functional form of the interface polarization is known in s-p coordinates. Thus, the rationale for using s-p coordinates is that, typically, the Jones matrix for a ray at an optical interface will be calculated in the s-p coordinates. Then it will be rotated into the system x-y coordinates. Once all the Jones matrices for the ray at all surfaces have been rotated into x-y coordinates, they can be cascaded to give the instrumental polarization along that ray path in the system x-y coordinates.

The s-p coordinates are based on the concept of the s and p planes. Consider light with unit wave vector \hat{k} (normalized to one) incident at a surface with normal \hat{n} . The plane of incidence, or "p-plane" is the plane which contains \hat{k} and \hat{n} . The plane perpendicular to the plane of incidence which contains \hat{k} is the "s-plane." Two unit vectors perpendicular to \hat{k} are defined to form an orthonormal basis for this coordinate system.

3.3 Polarization Decomposition

All optical elements display variation of their Jones matrices as the angle of incidence changes. Further, this always involves more than just a variation in the intensity and phase of the light; it also involves polarization and retardance. A fine optical element used in a transparent system will not display polarization effects at normal incidence; it may show some absorption, reflection loss, or phase shift, but not polarization or retardance.

The Jones matrix can be decomposed and can be expressed as

$$J = \sum_{k=0}^3 c(k) \sigma(k),$$

$$= c_0 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + c_1 \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} + c_2 \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + c_3 \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix},$$

where $\sigma(k)$ are the identity matrix and the Pauli spin matrices which describe the specific polarization state. The $c(k)$ define a vector which then describes the polarization properties of the element. At an angle of incidence i , the C vector will have the form

$$C_i = (c(0,i), c(1,i), c(2,i), c(3,i)) =$$

$$(\rho(0,i)e^{i\phi(0,i)}, \rho(1,i)e^{i\phi(1,i)}, \rho(2,i)e^{i\phi(2,i)}, \rho(3,i)e^{i\phi(3,i)})$$

where each component has an amplitude (ρ) and a phase (ϕ) part. The functional dependencies of the C vector coefficients are calculated from the Fresnel equations and coating equations for the interface.

3.4 Weak Polarizers

A weak polarization element is defined as a polarizer having a C vector such that for some range of i :

$$|c(0,i)|^2 \gg (|c(1,i)|^2 + |c(2,i)|^2 + |c(3,i)|^2)^{1/2}.$$

Such a polarization element transmits light in a polarization state similar to the incident state with only weak coupling into other polarization states. The polarization behavior is dominated by transmission with only traces of polarization or retardance. Any polarization present is at the few percent level or less, such that any incident linearly polarized beam has a transmission coefficient which varies a few percent or less with orientation. Similarly, the retardation is a few degrees or less, far less than a quarter-wave ($\pi/2$) retarder. Near normal incidence, metals in reflection (e.g., telescope mirrors), and dielectric refracting interfaces (e.g., relay lens) are weak polarizers. In addition, near normal incidence, antireflection coated lenses used in transmission and metals with reflection enhancing coatings are typically weak polarizers for wavelengths near the thin film design wavelength.

3.5 Amplitude Transmission Relations

The amplitude transmission equations for an interface are the equations which relate the amplitude and phase of the electric fields, \vec{E} , at an interface. The most general amplitude transmission equations for a non-scattering linear interface are:

$$\vec{E}'(s) = a(ss) \vec{E}(s) + a(ps) \vec{E}(p)$$

$$\vec{E}'(p) = a(sp) \vec{E}(s) + a(pp) \vec{E}(p)$$

where, for this section, the plane of incidence will be aligned with the y axis. This equation is equivalent to the Jones matrix equation,

$$\begin{bmatrix} \vec{E}'(s) \\ \vec{E}'(p) \end{bmatrix} = \begin{bmatrix} a(ss) & a(sp) \\ a(ps) & a(pp) \end{bmatrix} \begin{bmatrix} \vec{E}(s) \\ \vec{E}(p) \end{bmatrix}.$$

For interfaces whose eigenpolarizations are linear polarized, light-oriented parallel and perpendicular to the plane of incidence, the transfer of energy across the interface is separable into two uncoupled components which can be written in the form:

$$\vec{E}'(s) = a(s) \vec{E}(s) = \rho(s) \exp(i\phi(s)) \vec{E}(s)$$

$$\vec{E}'(p) = a(p) \vec{E}(p) = \rho(p) \exp(i\phi(p)) \vec{E}(p).$$

The amplitude transmission coefficients $a(s)$ and $a(p)$, or in polar coordinates, $\rho(s)$, $\phi(s)$, $\rho(p)$, and $\phi(p)$, are determined by the Fresnel equations for the interface. The type of energy transfer equation, where the s and p equations are separable, is a "separable amplitude transmission relation." Only polarization elements with linearly polarized light as the eigenpolarizations have the energy transfer equations in the separable form. This includes all the elements and coatings used in the SAMEX foreoptics.

The separable amplitude transmission relations correspond to a diagonal Jones matrix in s-p coordinates. The Jones matrix and C vector for an amplitude transmission interface in s-p coordinates are:

$$JJ(i) = \begin{bmatrix} a(s,i) & 0 \\ 0 & a(p,i) \end{bmatrix}$$

and

$$C = 1/2 [a(s,i)+a(p,i), a(s,i)-a(p,i), 0, 0].$$

3.6 Taylor Series Representation of SAMEX Coatings

In optical aberration theory, expressions for the optical path length of ray segments through the the optical system are obtained by performing a Taylor series expansion on Snells law, the law of reflection and the grating equation, to obtain expressions for the optical path length as a power series expansion in the ray coordinates. Thus Snells law,

$$n \sin i = n' \sin i'$$

is rewritten for i' as

$$i' = \arcsin[(n/n') \sin i]$$

or

$$i' = (n/n') i + [(n/n')^3 - (n/n')] i^3/6 + O(i^5).$$

The polarization aberrations are generated in an analogous fashion. To obtain the variation of the Jones matrix in the exit pupil of a system, the appropriate coating equations are required in Taylor series form. For radially symmetric optical systems, expansions in the angle of incidence about normal incidence are used.

An isotropic interface appears unchanged as it is rotated about the surface normal. Ideally, and to a first approximation, the SAMEX coatings are isotropic. So, for any isotropic interfaces the Fresnel equations are even functions since the surface does not distinguish between angles of incidence of $+i$ and $-i$.

An even function contains only even terms in its Taylor series expansion about the origin. Thus, if $f(i)$ is a reflection or transmission coefficient, the Taylor series representations of the coating equations have the form

$$f(i) = f_0 + f_2 i^2 + f_4 i^4 + \dots,$$

where

$$f_n = 1/n! \left. \frac{d^n f(i)}{di^n} \right|_{i=0}.$$

For weakly polarizing interfaces described by amplitude transmittance relations, the Taylor series forms of the Jones matrix and C vector are calculated as follows. First, the Taylor series expansion is determined for the amplitude transmission relations:

$$a(s,i) = a_0(s) + a_2(s) i^2 + a_4(s) i^4 + \dots,$$

$$a(p,i) = a_0(p) + a_2(p) i^2 + a_4(p) i^4 + \dots.$$

Then, the Taylor series expansion about $i = 0$ in s-p coordinates for the Jones matrix is

$$JJ(i) = \begin{bmatrix} a_0(s) & 0 \\ 0 & a_0(p) \end{bmatrix} + i^2 \begin{bmatrix} a_2(s) & 0 \\ 0 & a_2(p) \end{bmatrix} + \dots.$$

The corresponding C vector expansion in s-p coordinates is

$$C = [c(0,0) + c(0,2)i^2 + \dots, c(1,0) + c(1,2)i^2 + \dots, 0, 0]$$

where the n^{th} order C vector component is given by

$$\begin{aligned} c(0,n) &= 1/2 (a_n(s) + a_n(p)) \\ c(1,n) &= 1/2 (a_n(s) - a_n(p)). \end{aligned}$$

A matrix equation to calculate $c(0,n)$ and $c(1,n)$ from results of a thin film program is given by Chipman (1987). For the SAMEX coatings characterized by separable amplitude transmission relations, the diagonal and circular polarization components, $c(2,n)$ and $c(3,n)$, are always zero. The normalized C vector for the separable amplitude transmission relations is

$$C = \tau(q) [1 + d(0,2)i^2 + \dots, d(1,0) + d(1,2)i^2 + \dots, 0, 0]$$

where

$$r(q) = c(0,0) \text{ and } d(k,n) = c(k,n)/c(0,0) \quad ,$$

for the q^{th} optical interface.

The Jones matrix and C vector for coordinates other than the s-p coordinates are obtained from the polarization rotation operation. For example, the s-p coordinates are rotated with respect to the x-y coordinates by θ , the orientation of the plane of incidence. The Jones matrix in x-y coordinates $JJ(x,y)$ is related to the Jones matrix in s-p coordinates $JJ(sp)$ by the equation

$$JJ(x,y) = R(-\theta) JJ(s,p) R(\theta).$$

The Taylor series coefficients for the Fresnel equations which govern an uncoated dielectric or metal surface have been determined for use in determining the instrumental polarization of a conventional Cassegrain telescope. The notation $a(s)$ and $a(p)$ will refer to either the reflected or transmitted amplitude transmission coefficient, while $t(s)$, $t(p)$, $r(s)$, and $r(p)$ are used to refer unambiguously to the transmitted or reflected components. The Fresnel amplitude transmission equations are:

$$\begin{aligned} t(s,i) &= (2 \cos i \sin i')/(\sin(i+i')) \\ &= (2n \cos i)/(n \cos i + n' \cos i') \quad , \end{aligned}$$

$$\begin{aligned} t(p,i) &= (2 \cos i \sin i')/(\sin(i+i') \cos(i-i')) \\ &= (2n \cos i)/(n' \cos i + n \cos i') \quad , \end{aligned}$$

$$\begin{aligned} r(s,i) &= (-\sin(i-i'))/(\sin(i+i')) \\ &= (n \cos i - n' \cos i')/(n \cos i + n' \cos i') \quad , \end{aligned}$$

$$\begin{aligned} r(p,i) &= (\tan(i-i'))/(\tan(i+i')) \\ &= (n' \cos i - n \cos i')/(n' \cos i + n \cos i') \quad . \end{aligned}$$

The Fresnel equations depend on the ratio of the indices, n and n' , but not on the values of the refractive indices individually. This relative refractive index ratio is defined as

$$N = n/n' .$$

The Fresnel equations are equally valid for real n , corresponding to transparent media, or complex n , corresponding to absorbing media or metals.

The second-order Taylor series expansions for the Fresnel amplitude coefficients about $i = 0$ are:

$$\begin{aligned} t(s,i) &= (2N)/(N+1) + i^2 (N(N-1))/(N+1) , \\ t(p,i) &= (2N)/(N+1) + i^2 (N^2(N-1))/(N+1) , \\ r(s,i) &= (N-1)/(N+1) - i^2 (N-1)/(N(N+1)) , \\ r(p,i) &= (N-1)/(N+1) + i^2 (N-1)/(N(N+1)) . \end{aligned}$$

The direct method for calculating the Taylor series coefficients of a coating series given in the last section are impractical for multi-layer coatings due to the complexity of calculating the partial derivatives of the appropriate amplitude transmission equations. The Taylor series coefficients can be obtained numerically from the s and p amplitude transmissions evaluated at a series of angles of incidence. An algorithm to sixth order has been given by Chipman (1987). The algorithm was used with the thin film design programs Filmstar and Films to obtain the Taylor series expansions of the transmitted and reflective amplitudes as a function of the angle of incidence for use in the polarization aberration calculations for SAMEX.

3.7 Cascaded Weak Polarizers

In this section the Jones matrix describing the instrumental polarization for light propagating along a ray path through the SAMEX foreoptics is derived. Results are also given for the instrumental polarization associated with paraxial rays as functions of the Taylor series of the C vectors representing the optical interfaces. The notation used in this section is compiled in Table D.3.

Consider an optical system with Q optical interfaces numbered in the order encountered from $q = 1$ to Q . No symmetry regarding the optical configuration is assumed. Light propagates along a specified ray path such as would be calculated by an optical ray trace calculation. At each interface some polarization is introduced due to differences in the optical constants across the interface. In addition, polarization is associated with the ray path between interfaces due to optically active crystals, dichroism, birefringence, gradient index

Table D.3. Notation for section.

C	C vector
d(k)	Normalized C vector components in s-p coordinates
c(k)	d(k) coefficients rotated into arbitrary plane of incidence
i	Angle of incidence
JJ	Jones matrix
k	Pauli spin matrix index: 0,1,2,3
d	Length of a ray segment
q	Surface index
Q	Total number of surfaces
v	Orientation of the plane of incidence
r(k)	Absorption or polarization coefficient
σ	Pauli spin matrix
τ	Normal transmittance
p(k)	Phase or retardance coefficient

Subscript Ordering: k, 2, q.

For example, d(1,2,3) is the coefficient for the $\sigma(1)$ polarization basis state that is second order in the angle of incidence Taylor series, i^2 , for q = 3, the third interface.

materials, or other polarizing mechanisms. But for the polarization analysis for SAMEX foreoptics, polarization associated with the optical path between interfaces was zero. Therefore only the weak interface-induced polarization is considered here.

3.7.1 Homogeneous optical systems

A homogeneous interface has optical properties independent of spatial coordinates on the interface. The Jones matrices are functions only of the angle of incidence, plane of incidence, and optical properties of the interface media, $JJ = JJ(i, \theta, n, n')$ and similarly $C = C(i, \theta, n, n')$. The foreoptics and polarimeter sections of the SAMEX magnetograph are homogeneous optical systems.

Likewise, a homogeneous medium has optical properties independent of spatial coordinates. An anisotropic crystalline medium is homogeneous if it consists of a single crystal. The refractive index varies with direction but not with position.

3.7.2 Radially symmetric systems of lenses, mirrors, and coatings

The polarization properties of optical systems comprised of lenses, mirrors, and "fine" coatings will be developed. A

radially symmetric optical system has an axis of symmetry, the optical axis. It is assumed that the optical elements and materials used in transmission are highly transparent and non-polarizing, as is usual in lenses. The polarization contribution from the path lengths through highly transparent elements is small relative to the polarization arising at the interfaces and is neglected.

The polarization associated with ray paths near the optical axis, or in the paraxial regime, will be derived. For this paraxial development to be accurate, it is only necessary that the angles of incidence are small enough that the polarization associated with the interfaces is adequately approximated by a second-order expansion of the C vector as a function of the angle of incidence. For an uncoated lens or mirror, this approximation is generally valid for $i < 30^\circ$. Calculation of the fourth- and higher-order coefficients allows estimation of the accuracy of these second-order equations. The paraxial region for this polarization analysis is typically orders of magnitude larger than the paraxial region of geometrical optics (the region where the fourth- and higher-order wave front aberrations are negligible.)

Homogeneous and isotropic interfaces do not display polarization at normal incidence. There is only an amplitude and phase change which is represented by the complex number, $\tau(q)$, the normal amplitude transmittance. An isotropic interface such as a lens, mirror, or coating has a C vector Taylor series in s-p coordinates ($\theta = 0$) of the form

$$C(i, \theta) \big|_{\theta=0} = \tau(q) [1, 0, 0, 0] + i^2 \tau(q) [d(0, 2), d(1, 2), 0, 0] \\ + i^4 \tau(q) [d(0, 4), d(1, 4), 0, 0] + \dots$$

For an arbitrary orientation θ of the plane of incidence, the C vector is

$$C(i, \theta) = \tau(q) [1, 0, 0, 0] + i^2 \tau^2(q) [c(0, 2), c(1, 2), c(2, 2), 0] \\ + i^4 \tau^2(q) [c(0, 4), c(1, 4), c(2, 4), 0] + \dots$$

where the c's are determined from the d's by a rotational change of basis. Since homogeneous and isotropic interfaces do not display circular retardance or circular polarization, $\sigma(3)$ is not included to simplify the mathematics.

3.7.3 The C vector for a paraxial ray

The SAMEX instrumental polarization will be analyzed by a paraxial optics development. Consider a paraxial ray path through an optical system from surfaces $q = 1$ to Q with angles of

incidence $i(q)$, and orientations of the plane of incidence $\theta(q)$. The Jones vector associated with the axial ray (down the optical axis) $i(q) = 0$ for all q , is

$$JJ = \tau(q) \sigma(0) \tau(q-1) \sigma(0) \dots \tau(2) \sigma(0) \tau(1) \sigma(0) = \tau \sigma(0)$$

where now

$$\tau \equiv \prod_{q=1}^Q \tau(q) .$$

The complex amplitude transmittance down the axis, τ , is the product of the normal incidence complex amplitude transmittances at each surface.

The Jones matrix associated with a ray at interface q can be expressed in terms of the expansion of the interface Jones matrix as

$$JJ(i,v,q) = \tau(q) [\sigma(0) + i^2(q)(c(0,2,q)\sigma(0) + c(1,2,q)\sigma(1) + c(2,2,q)\sigma(2)) \\ + i^4(q)(c(0,4,q)\sigma(0) + c(1,4,q)\sigma(1) + c(2,4,q)\sigma(2)) + \dots] .$$

The Jones matrix associated with the entire paraxial ray path resulting from keeping terms to second order at each interface is (where X represents multiplication carried onto the next line)

$$JJ = \tau(q)(\sigma(0) + i^2(q)(c(0,2,q)\sigma(0) + c(1,2,q)\sigma(1) + c(2,2,q)\sigma(2))) X \\ \tau(q-1)(\sigma(0) + i^2(q-1)(c(0,2,q-1)\sigma(0) + c(1,2,q-1)\sigma(1) + c(2,2,q-1)\sigma(2))) X \\ \dots X \\ \tau(2)(\sigma(0) + i^2(2)(c(0,2,2)\sigma(0) + c(1,2,2)\sigma(1) + c(2,2,2)\sigma(2))) X \\ \tau(1)(\sigma(0) + i^2(1)(c(0,2,1)\sigma(0) + c(1,2,1)\sigma(1) + c(2,2,1)\sigma(2))) .$$

Associated with each interface are four terms. Carrying out all the multiplications leads to 4^Q terms, all in even powers in i . Collecting terms of equal power in i , there is one term at zero'th order and $3Q$ terms at second order. If i is assumed small, the large number of higher-order terms is of diminishing importance. Collecting zero- and second-order terms in JJ_0 and JJ_2 , the expression for JJ is

$$\begin{aligned}
JJ_0 + JJ_2 = & \tau \sigma(0) + \tau \sigma(0) \sum_{q=1}^Q i^2(q) c(0,2,q) \\
& + \tau \sigma(1) \sum_{q=1}^Q i^2(q) c(1,2,q) \\
& + \tau \sigma(2) \sum_{q=1}^Q i^2(q) c(2,2,q).
\end{aligned}$$

Since no polarization or retardance was assumed on axis, the contributions to the second-order polarization for this ray are just sums of contributions from each surface. The multiplication taking place at second order for the $\sigma(1)$ term is of the form

$$\begin{aligned}
& i^2(q) \sigma(1) \sigma(0) \sigma(0) \sigma(0) \dots \sigma(0) + \sigma(0) i^2(q-1) \sigma(1) \sigma(0) \sigma(0) \dots \sigma(0) + \\
& \dots + \sigma(0) \sigma(0) \sigma(0) \sigma(0) \dots i^2(q) \sigma(1)
\end{aligned}$$

where the c -dependence is not shown explicitly.

This equation contains the useful result that, when no elements display polarization or retardance at normal incidence, as in the SAMEX foreoptics, there is no order dependence in the second-order terms. Only one non-identity matrix term occurs in each second-order matrix product. The second-order polarization associated with the paraxial ray path is obtained by a simple summation of second-order polarization contributions at each intercept. Chipman (1987) gives a complete account of this derivation.

3.8 Paraxial Optics Geometry

The polarization aberrations for SAMEX are a description of the polarization behavior of an optical system expressed as an expansion about the center of the object and the center of the pupil. Thus it is appropriate and convenient to obtain the derivations from a paraxial ray trace; appropriate, because understanding the instrumental polarization near the center of the pupil and image is key to understanding instrumental polarization in general; convenient because the paraxial ray trace is linear, and thus easy to manipulate.

The paraxial coordinate system used is a normalized right-handed coordinate system. The z axis is the optical axis of a rotationally symmetric optical system. Light initially travels in the direction of increasing z . Figure D.1 shows the notation.

For a rotationally symmetric system, the object can be located on the y axis without loss of generality. The object coordinate H is normalized such that $H = 0$ in the center of the field (on the optical axis) and $H = 1$ at the nominal edge of the field of view.

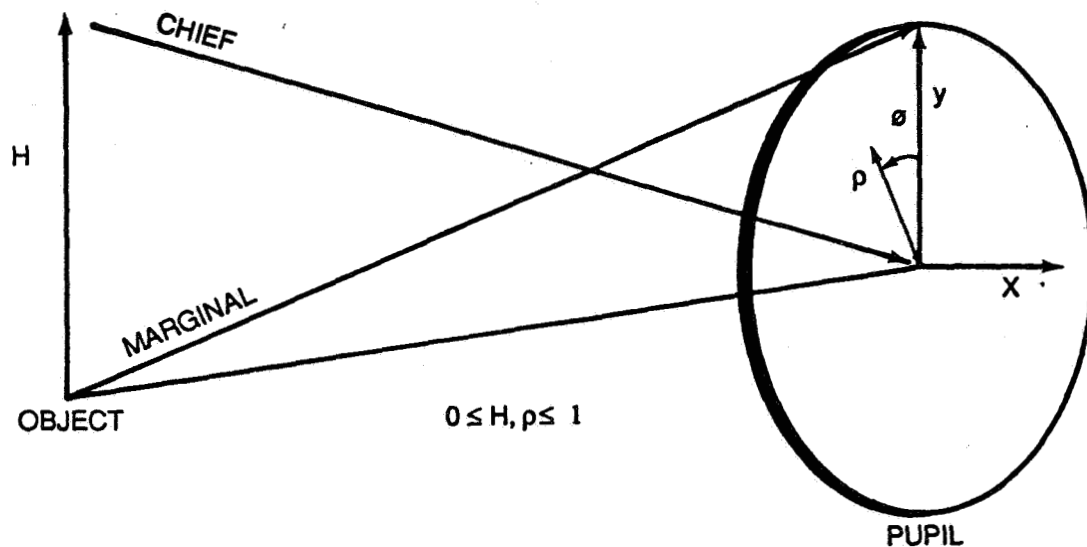


Figure D.1. Paraxial coordinate system. The paraxial system is a normalized right-handed coordinate system. The z axis is the optical axis of a rotationally symmetric optical system; light initially travels in the direction of increasing z . Rays through an optical system are characterized by ray coordinates at the object and entrance pupil. H is the normalized object coordinate, ρ is the normalized pupil radius, and ϕ is the polar angle in the pupil measured counterclockwise from the y axis. The normalized Cartesian coordinates in the pupil are x and y . The chief and marginal rays are also shown.

The location where a ray strikes the entrance pupil is specified by the polar pupil coordinates ρ and ϕ . ρ is normalized such that at the edge of a circular pupil $\rho = 1$. The angle ϕ is defined here as it is in much of geometric optics, and in defiance to most analytical geometry, as being zero on the y axis and increasing counterclockwise. Normalized Cartesian pupil coordinates x and y can be used. They are defined as: $x = \rho \sin(\phi)$ and $y = \rho \cos(\phi)$.

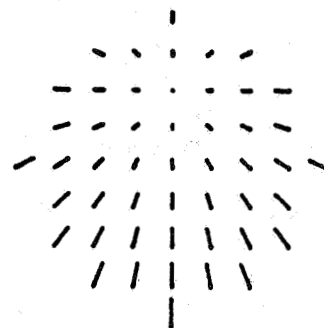
Expressions for the angle of incidence i and the orientation of the plane of incidence θ of a ray at a given surface q will be expressed in terms of the angles of incidence of the marginal $i(m,q)$ and chief $i(c,q)$ rays at that surface. Details of the derivation may be found in Chipman (1987, section D). Note, however, for a radially symmetric system the angle of incidence should be a function of H^2 , ρ^2 , and $H\rho \cos\phi$ since the function should be invariant to rotation of the system about the optical axis and must give the same result when $-x$ is substituted for $+x$ where $x = \rho \cos\phi$.

Assume that a paraxial ray trace has been performed for a specific system and that $i(m,q)$ and $i(c,q)$ have been calculated. A ray from normalized object coordinate H which passes through pupil coordinates ρ and ϕ will have an angle of incidence $i(q)$ and orientation of the plane of incidence $\theta(q)$ at surface q equal to:

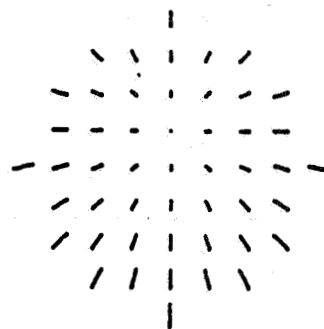
$$i(q) = [H^2 i^2(c,q) + 2H\rho \cos(\phi) i(c,q) i(m,q) + \rho^2 i^2(m,q)]^{1/2}$$

$$\theta = \sin^{-1} [\rho \sin(\phi) i(m)/|i|] .$$

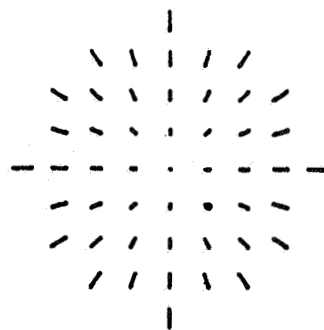
Figure D.2 shows the paraxial angle and plane of incidence for three field angles. The magnitude of the angle of incidence is represented by the length of the lines. The orientation of the lines corresponds to the orientation of the plane of incidence. Note that for off axis, the pattern is a shifted version of the on-axis pattern. The incidence angle is given by the pupil coordinate (ρ, ϕ) and the image coordinate (H) since these coordinate define the optical path of a single ray.



FIELD ANGLE, $2H_0$



FIELD ANGLE, H_0



FIELD ANGLE, ON AXIS

Figure D.2. Paraxial angles and plane of incidence for three field angles. The magnitude of the angle of incidence is represented by the length of the lines. The orientation of the lines corresponds to the orientation of the plane of incidence. Note that in the off-axis cases, the pattern is a shifted version of the on-axis pattern.

4. Polarization Aberrations

This section derives the polarization aberrations for SAMEX as a Taylor series description of the instrumental polarization associated with paraxial rays through the foreoptics. Polarization aberration is a description of the polarization behavior of an optical system expressed by an expansion about the center of the object and center of the pupil. Table D.4 gives an overview of polarization aberration theory. The results are obtained in a form very similar to the wave front aberrations. In particular, terms closely related to defocus, tilt, piston error as well as the Seidel and higher-order aberrations can be associated with all eight of the basis Jones matrices. Since polarization effects are typically orders of magnitude smaller than wave front effects, fewer terms are needed for a sufficient description. A method of calculating aberration coefficients for specific systems is developed in the next section.

4.1 The Polarization Aberration Expansion

The wave front polynomial expansion describes the variation of the optical path difference through an optical system as a function of ray coordinates. A closely related expansion will be presented for all four basis polarization matrices $\sigma(0)$, $\sigma(1)$, $\sigma(2)$, $\sigma(3)$. The polarization aberration expansion for radially symmetric systems uses a very similar polynomial expansion to describe all eight basis polarization vectors. The principal difference is a modified form for the linear polarization and linear retardance terms since these involve both a magnitude and an orientation.

The eight forms of polarization behavior can be characterized by four complex numbers, for example, the four elements of either the Jones matrix or the C vector. We introduce a new set of complex parameters, the complex polarization aberration coefficients, which gives a description of the polarization behavior of an optical interface. It should be emphasized that the amplitude and phase of the coefficients are generally unrelated. They refer to different aspects of the instrumental polarization. The amplitude part of the coefficient describes amplitude and polarization effects while the phase part describes phase and retardance. When necessary the amplitude and phase of the polarization aberration coefficient, P , will be denoted by A and ϕ where,

$$P = A \exp(i \phi).$$

Table D.4. Polarization Aberration Theory Outline

(1) Jones Vector

The Jones vector is a complex two-component vector describing the electromagnetic field.

$$\vec{J} = \begin{bmatrix} E_x \\ E_y \end{bmatrix}$$

(2) Paraxial Ray Trace

The angle of incidence, i , of any ray can be written in terms of the angle of incidence of the chief and marginal rays, $i(c)$ and $i(m)$.

$$i(c), i(m)$$

(3) Coating Calculation

Thin film design programs calculate the amplitude transmission coefficients for s and p light, $t(s,i)$ and $t(p,i)$.

$$t(s,i), t(p,i)$$

(4) Taylor Series Representation

The amplitude transmission coefficients, $t(s,i)$ and $t(p,i)$, are transformed to a Taylor series expansion about normal incidence.

$$t(s,i) = t_0(1 + t(s,2)i^2 + \dots)$$

$$t(p,i) = t_0(1 + t(p,2)i^2 + \dots)$$

(5) Jones Matrix for an Interface

The second-order Jones matrix for an interface can be written in terms of three polarization aberration terms represented by coefficients.

$$P(1,0,2,2,q), P(1,1,1,1,q),$$

$$P(1,2,0,0,q)$$

(6) Aberrations Sum for System

The second-order polarization aberrations for the system is the sum of the aberration contributions of each interface.

$$P(1,0,2,2) = \sum_q P(1,0,2,2,q)$$

$$P(1,1,1,1) = \sum_q P(1,1,1,1,q)$$

$$P(1,2,0,0) = \sum_q P(1,2,0,0,q)$$

(7) Polarization Accuracy

Summarize the performance of the system with a single number, the polarization accuracy (Δ_p). This is the maximum fraction of light coupled into the orthogonal polarization state. This occurs at the edge of the field of view.

$$\Delta_p$$

The complex aberration coefficient is written to contain polarization, amplitude effects, retardance, and phase effects in a single term. The following polarization aberration expansion for zeroth and second order is used. The subscripts are defined as follows:

$$P(k,u,v,w) = A(k,u,v,w) \exp(i\phi(k,u,v,w)),$$

where: k is the type of polarization behavior,
 u is the order of the H dependence, H^u ,
 v is the order of the ρ dependence, ρ^v , and,
 w is the order of the ϕ dependence, $(\cos(\phi))^w$.

The indices u , v , and w are used exactly as they are for the wave front aberrations, as shown in the next section.

The polarization aberration expansion of the Jones matrix for the SAMEX foreoptics is

$$\begin{aligned} JJ(H,\rho,\phi) &= \sum_{k=0}^3 c(k,H,\rho,\phi) \sigma(k) \\ &= \sum_{k=0}^3 \sum_u \sum_v \sum_w P(k,u,v,w) H^u \rho^v \cos^w(\phi) \sigma(k). \end{aligned}$$

Here we define the aberration coefficients as an expansion of the Jones matrix in terms of the ray coordinates ρ , ϕ , and H . The C vector coefficients in this polarization aberration expansion are:

amplitude and phase term:

$$\begin{aligned} c(0,H,\rho,\phi) &= A(0,0,0,0) + A(0,2,0,0)H^2 \\ &\quad + A(0,1,1,1) H \rho \cos(\phi) + A(0,0,2,0) \rho^2 \\ &\quad + \gamma(\phi(0,0,0,0) + \phi(0,2,0,0)H^2 \\ &\quad + \phi(0,1,1,1)H \rho \cos(\phi) + \phi(0,0,2,0) \rho^2), \end{aligned}$$

linear polarization and retardance terms:

$$\begin{aligned}
c(1,H,\rho,\phi) = & A(1,0,0,0) + A(1,2,0,0) H^2 \\
& + H \rho (A(1,1,1,1) \cos(\phi) - A(2,1,1,1) \sin(\phi)) \\
& + \rho^2 (A(1,0,2,0) \cos(2\phi) - A(2,0,2,0) \sin(2\phi)) \\
& + \tilde{\gamma} (\Phi(1,0,0,0) + \Phi(1,2,0,0) H^2 + \\
& H \rho (\Phi(1,1,1,1) \cos(\phi) - \Phi(2,1,1,1) \sin(\phi)) + \\
& \rho^2 (\Phi(1,0,2,0) \cos(2\phi) - \Phi(2,0,2,0) \sin(2\phi))).
\end{aligned}$$

diagonal polarization and retardance terms:

$$\begin{aligned}
c(2,H,\rho,\phi) = & A(2,0,0,0) + A(2,2,0,0) H^2 \\
& + H \rho (A(2,1,1,1) \cos(\phi) + A(1,1,1,1) \sin(\phi)) \\
& + \rho^2 (A(2,0,2,0) \cos(2\phi) + A(1,0,2,0) \sin(2\phi)) \\
& + \tilde{\gamma} (\Phi(2,0,0,0) + \Phi(2,2,0,0) H^2 + \\
& H \rho (\Phi(2,1,1,1) \cos(\phi) + \Phi(1,1,1,1) \sin(\phi)) + \\
& \rho^2 (\Phi(2,0,2,0) \cos(2\phi) - \Phi(1,0,2,0) \sin(2\phi))).
\end{aligned}$$

circular polarization and retardance terms:

$$\begin{aligned}
c(3,H,\rho,\phi) = & A(3,0,0,0) + A(3,2,0,0) H^2 + \\
& + A(3,1,1,1) H \rho \cos(\phi) + A(3,0,2,0) \rho^2 \\
& + \tilde{\gamma} (\Phi(3,0,0,0) + \Phi(3,2,0,0) H^2 \\
& + \Phi(3,1,1,1) H \rho \cos(\phi) + \Phi(3,0,2,0) \rho^2).
\end{aligned}$$

There are thirty-two terms in this polarization aberration expansion to second order arising from four terms in each of the eight degrees of freedom of the Jones matrix. The terms may be grouped as follows:

A(0,u,v,w)	Amplitude terms
A(1,u,v,w)	Linear polarization terms
A(2,u,v,w)	Diagonal polarization terms
A(3,u,v,w)	Circular polarization terms
$\Phi(0,u,v,w)$	Wave front or phase terms

$\Phi(1,u,v,w)$	Linear retardance terms
$\Phi(2,u,v,w)$	Diagonal retardance terms
$\Phi(3,u,v,w)$	Circular retardance terms
$P(k,0,0,0)$	"Constant Piston" terms
$P(k,2,0,0)$	"Quadratic Piston" terms
$P(k,1,1,1)$	"Tilt" terms
$P(k,0,2,0)$	"Scalar Defocus" terms
$P(k,0,2,2)$	"Vector Defocus" terms

The names of the wave front aberrations: piston, quadratic piston, defocus and tilt, are used here, in an extended sense, to describe variations of components of the Jones vector which share the same functional dependencies as the wave front aberrations. Defocus is a ρ^2 variation of a parameter. Tilt is an $H \rho \cos(\phi)$ variation. Quadratic piston is an H^2 variation. Thus, "amplitude defocus" means a ρ^2 amplitude variation. Likewise the "circular retardance tilt" is the $H \rho \cos(\phi)$ circular retardance variation, and so on.

This polarization aberration expansion is an equation which describes all possible second-order variations of the Jones matrix, just as the second-order wave front aberration expansion spans the set of all second-order wave front variations. Thus, the polarization aberration expansion characterizes quadratic variations of all forms of wave front, amplitude, polarization, and retardance.

This polarization aberration expansion is a summation of terms in the different Pauli spin matrix components, not a product. Thus, the four C vector elements can be pictured as acting in parallel, almost side by side in the aperture, but not in series. Each term describes an amount of a particular form of polarization, independent of the other contributions.

An "aberration term" is to be considered as containing all the algebraic terms in the expansion with the same coefficient. Most of the coefficients occur only once and the aberration term contains only one algebraic term. The exceptions are the terms, $A(1,1,1,1)$, $\Phi(1,1,1,1)$, $A(1,0,2,2)$, $\Phi(1,0,2,2)$, $A(2,1,1,1)$, $\Phi(2,1,1,1)$, $A(2,0,2,2)$, and $\Phi(2,0,2,2)$. These aberration terms have components both along the axes and at 45° .

With SAMEX, the principal concerns are with the linear piston, linear tilt, and linear defocus terms, both in polarization and retardance. These are going to be the largest terms present which corrupt the incident polarization state. These values are given in Table 19 in the report for the SAMEX optics.

For a detailed discussion of the physical meaning of the polarization aberration coefficients, see Chipman (1987). However, a discussion of the origin of tilt and piston terms, $P(0,2,0)$, $P(1,1,1)$, and $P(2,0,0)$, is included below.

A distinction is made between scalar and vector aberrations. The wave front aberrations are scalar aberrations, single-valued functions of object and pupil coordinates. The linear polarization and linear retardance aberrations are vector aberrations since a magnitude and orientation is associated with these at each point. Amplitude, circular polarization, and circular retardance aberrations are scalar since they are single valued and range positive and negative.

Figure D.3 (top) shows the chief and limiting rays at an interface for objects on axis and at the edge of the field of view. Figure D.3 (bottom) is a plot of the value of the angle of incidence along the y axis as a function of ρ . Tilt terms naturally occur because as the object point moves off axis, the angle of incidence increases at one edge of the beam and decreases at the other edge. Tilt contains the first-order portion of this correction.

Figure D.4 shows the off-axis angle of incidence squared and the decomposition of this into defocus, tilt, and piston terms. These terms are required to describe a quadratic variation whose vertex is located at an arbitrary position on the y axis because

$$x^2 + (y-a)^2 = x^2 + y^2 - 2y a + a^2 .$$

In this case since a is a linear function of H ,

$$a = k H ,$$

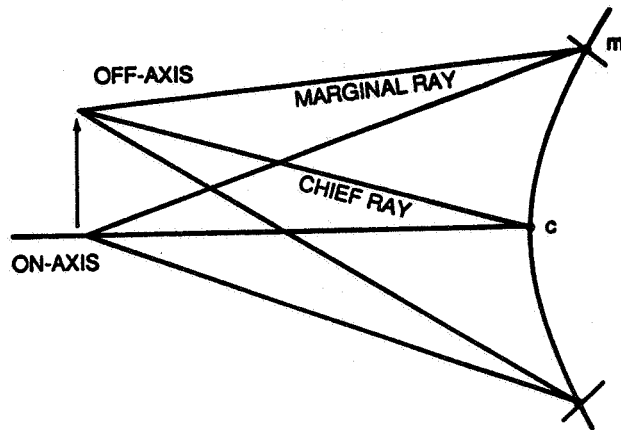
then the quadratic polarization variation becomes

$$\begin{aligned} c(x^2 + (y-kH)^2) &= c(x^2+y^2) - 2c y k H + c k^2 H^2 \\ &= P(0,2,0)\rho^2 + P(1,1,1)2H \rho \cos(\phi) + P(2,0,0)H^2, \end{aligned}$$

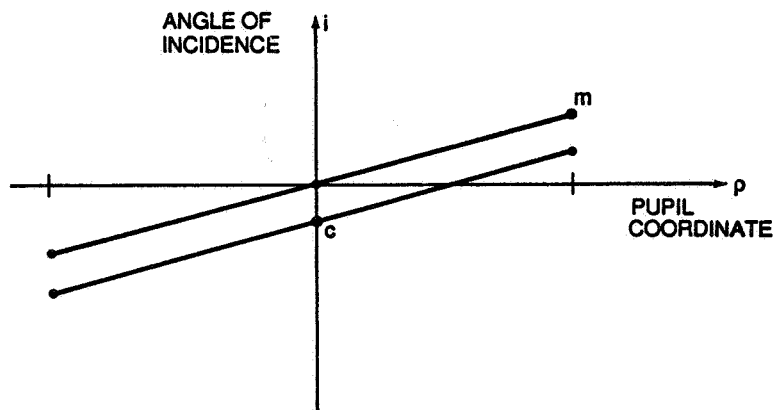
where $P(0,2,0)$, $P(1,1,1)$, and $P(2,0,0)$ are the defocus, tilt, and quadratic piston aberration coefficients and the polarization index is not shown. Tilt and piston terms arise naturally from decentered defocus. Similarly, the fourth-order wave front aberrations coma, astigmatism, field curvature, and distortion arise naturally from decentered spherical aberration, ρ^4 .

4.2 Calculation of Aberration Coefficients

The method used to calculate the second-order polarization aberration coefficients for the SAMEX foreoptics given the C vector power series for each interface is detailed in this section.



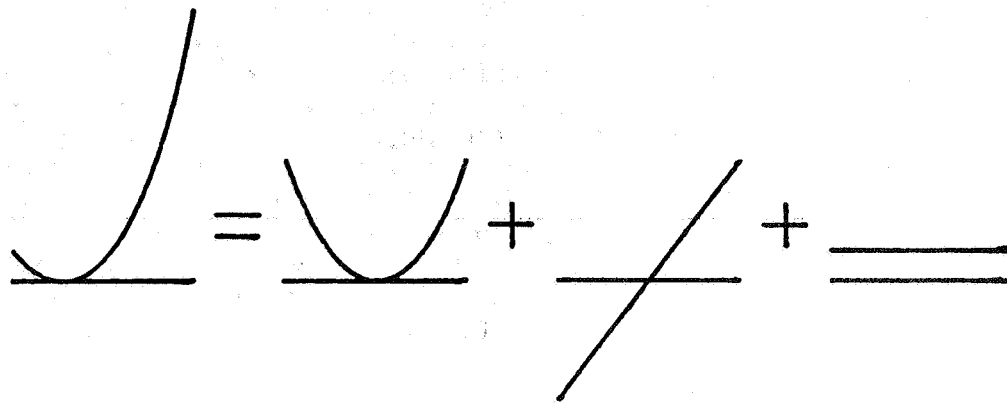
A.



B.

Figure D.3. Angles of incidence for objects on and off axis. In the top frame, the chief and limiting rays at an interface are shown for objects on the optical axis and at the edge of the field of view. In the lower part of the figure, the angle of incidence for rays incident along the y axis (in the paraxial system) is plotted as a function of the normalized radius ρ .

C-5



$$(x+a)^2 = x^2 + 2ax + a^2$$

Off-axis Angle of Incidence Squared	Defocus	Tilt	Piston
--	---------	------	--------

Figure D.4. Quadratic effects of off-axis angles of incidence. The square of the off-axis angle of incidence is shown along with its decomposition into defocus, tilt, and piston terms.

4.2.1 Single surface aberrations for amplitude transmittance relations

For homogeneous and isotropic interfaces characterized by amplitude transmittance relations, such as lenses, mirrors, and thin film coatings, the polarization aberrations at an interface simplify considerably. At these interfaces the Fresnel equations and related thin film equations are separable into s and p components, so the Jones matrices representing the interface in s-p coordinates are diagonal. The off-diagonal terms, diagonal polarization $\sigma(2)$, and circular polarization $\sigma(3)$ are not present. Further, with isotropic media, the s and p amplitude transmission coefficients at normal incidence must be equal. Thus, the amplitude transmission functions can be expanded as:

$$\begin{aligned} a(s,i) &= a(0)(1 + a(s,2)i^2 + \dots) \exp(\tilde{\gamma}(d(0) + d(s,2)i^2 + \dots)) \\ &\approx \tau (1 + (a(s,2) + \tilde{\gamma} d(s,2))i^2 + \dots), \end{aligned}$$

$$\begin{aligned} a(p,i) &= a(0)(1 + a(p,2)i^2 + \dots) \exp(\tilde{\gamma}(d(0) + d(p,2)i^2 + \dots)) \\ &\approx \tau (1 + (a(p,2) + \tilde{\gamma} d(p,2))i^2 + \dots), \end{aligned}$$

where:

$$\tau = a(0) \exp(\tilde{\gamma} d(0)).$$

The s-p coordinate Jones matrix expansion to second order is

$$\begin{aligned} JJ(i) &= \begin{bmatrix} a(s,i) & 0 \\ 0 & a(p,i) \end{bmatrix} \\ &= \tau (\sigma(0)(1 + A(2) + \tilde{\gamma}D(2))i^2 + \sigma(1)(a(2) + \tilde{\gamma}d(2))i^2), \end{aligned}$$

where

$$A(2) = (a(s,2) + a(p,2))/2,$$

$$a(2) = (a(s,2) - a(p,2))/2,$$

$$D(2) = (d(s,2) + d(p,2))/2,$$

$$d(2) = (d(s,2) - d(p,2))/2.$$

The s-p coordinate, C vector Taylor series expansion to second order is

$$C(i) = \tau(1,0,0,0) + i^2 \tau(A(2)+\tilde{D}(2), a(2)+\tilde{d}(2), 0, 0).$$

The x-y coordinate, C vector Taylor series for orientation of the plane of incidence θ is

$$C(i, \theta) = \tau(1,0,0,0) + i^2 \tau(A(2)+\tilde{D}(2), (a(2)+\tilde{d}(2))\cos(2\theta), (a(2)+\tilde{d}(2))\sin(2\theta), 0).$$

The normal-incidence polarization aberration terms (the constant piston terms) are zero:

$$P(1,0,0,0) = P(2,0,0,0) = P(3,0,0,0) = 0.$$

There is no polarization or retardance on axis, only the amplitude and phase transmission factor τ .

All terms for off-axis diagonal and circular polarization are zero:

$$P(2,u,v,w) = P(3,u,v,w) = 0.$$

Thus, the single surface C vector in paraxial coordinates is obtained by substituting the paraxial representation of $i(H, \rho, \phi)$ and $v(H, \rho, \phi)$ into $C(i, v)$ yielding

$$c(0, H, \rho, \phi) = \tau + \tau d(0, 2) [H^2 i^2(c) + 2H\rho \cos(\phi) i(c)i(m) + \rho^2 i^2(m)],$$

$$c(1, H, \rho, \phi) = \tau d(1, 2) [H^2 i^2(c) + 2H\rho \cos(\phi) i(c)i(m) + \rho^2 \cos(2\phi) i^2(m)],$$

$$c(2, H, \rho, \phi) = \tau d(1, 2) [2H\rho \sin(\phi) i(c)i(m) + \rho^2 \sin(2\phi) i^2(m)],$$

$$c(3, H, \rho, \phi) = 0.$$

Since there is no diagonal polarization, the only contributions to $c(2)$ arises from the rotation of linear polarization from the s-p coordinates into the x-y coordinates.

4.2.2 Polarization aberration coefficients for systems

Since the polarization aberrations are only being evaluated to second order in the angle of incidence, the difference between spheres, parabolas, conics, or other radially symmetric aspherics does not occur at this order. The relevant shape parameter here is only the vertex radius of curvature. The angle and plane of incidence differences between these types of interfaces are the same at second order but will differ at fourth order and higher.

For surfaces $q = 1$ to Q , each surface is characterized by three complex parameters from the normalized C vector expansion:

$$d(0,0) = \tau(q), d(0,2,q), \text{ and } d(1,2,q) .$$

The single surface polarization aberration coefficients are:

$$\begin{aligned} P(0,0,0,0,q) &= \tau(q), \\ P(0,2,0,0,q) &= \tau(q) d(0,2,q) i^2(c), \\ P(0,1,1,1,q) &= 2\tau(q) d(0,2,q) i(c)i(m), \\ P(0,0,2,0,q) &= \tau(q) d(0,2,q) i^2(m), \\ P(1,2,0,0,q) &= \tau(q) d(1,2,q) i^2(c), \\ P(1,1,1,1,q) &= 2\tau(q) d(1,2,q) i(c)i(m), \\ P(1,0,2,2,q) &= \tau(q) d(1,2,q) i^2(m). \end{aligned}$$

The polarization aberration coefficients for the system are calculated by chain multiplying the single surface polarization aberration expressions and keeping terms to second order in H and ρ . The zero- and second-order Jones matrices for the q^{th} interface are:

$$\begin{aligned} JJ_0(q,H,\rho,\phi) &= P(0,0,0,0,q) \sigma(0) = d(0,0,q) \sigma(0) = \tau(q) \sigma(0), \\ JJ_2(q,H,\rho,\phi) &= \sigma(0) [H^2 P(0,2,0,0,q) + 2H\rho \cos(\phi)P(0,1,1,1,q) \\ &\quad + \rho^2 P(0,0,2,0,q)] \\ &\quad + \sigma(1) [H^2 P(1,2,0,0,q) + 2H\rho \cos(\phi)P(1,1,1,1,q) \\ &\quad + \rho^2 \cos(2\phi)P(1,0,2,2,q)] \\ &\quad + \sigma(2) [2H\rho \sin(\phi)P(1,1,1,1,q) + \rho^2 \sin(2\phi)P(1,0,2,2,q)] \\ &= \sigma(0)\tau(q)d(0,2) [H^2 i^2(c) + 2H\rho \cos(\phi)i(c)i(m) + \rho^2 i(m)^2] \end{aligned}$$

$$\begin{aligned}
& + \sigma(1)\tau(q)d(1,2)[H^2 i(c)^2 + 2H\rho \cos(\phi)i(c)i(m) + \rho^2 \cos(2\phi)i^2(m)] \\
& + \sigma(2)\tau(q)d(1,2)[2H\rho \sin(\phi)i(c)i(m) + \rho^2 \sin(2\phi)i^2(m)].
\end{aligned}$$

Multiplication of the single surface Jones matrices yields

$$\begin{aligned}
JJ(H,r,p) &= \prod_{q=Q,-1}^{\Pi} JJ(q,H,\rho,\phi) \\
&= \prod_{q=Q,-1}^{\Pi} [JJ_0(q) + JJ_2(q,H,\rho,\phi)].
\end{aligned}$$

Since $JJ_0(q)$ is a constant function, independent of H , ρ , and ϕ , the (H,ρ,ϕ) dependence can be dropped. This expression contains 2^Q terms. The order of a term is the sum of the powers of H and ρ , $H^u \rho^v \cos(\phi)^w$, i.e., order = $u + v$. There is one first-order term and Q second-order terms.

The zero-order Jones matrix is

$$JJ_0 = \prod_{q=Q,-1}^{\Pi} JJ_0(q,H,\rho,\phi) = \prod_{q=1}^{\Pi} \tau(q) = \tau,$$

and τ is now the system amplitude transmittance. The second-order Jones matrix is greatly simplified since, for isotropic surfaces, all zeroth-order Jones matrices are a constant times the identity matrix $\sigma(0)$. The second order only contains products which contain a single second-order term. The second-order Jones matrix is

$$JJ_2(H,\rho,\phi) = \sum_{q=1,Q}^{\Sigma} JJ_2(q,H,\rho,\phi).$$

At second order the weakly polarizing isotropic interfaces do not display order dependence. The product of any two second-order terms is fourth order. The order dependence enters at fourth and higher order. Second order is a simple sum of polarization contributions. Collecting the piston, tilt, and defocus terms from the second-order Jones matrix provides the coefficients for the system polarization aberration expansion to second order:

$$\begin{aligned}
P(0,0,0,0) &= \tau, \\
P(0,2,0,0) &= \tau \sum_{q=1,Q} d(0,2,q) i^2(c), \\
P(0,1,1,1) &= 2\tau \sum_{q=1,Q} d(0,2,q) i(c)i(m), \\
P(0,0,2,0) &= \tau \sum_{q=1,Q} d(0,2,q) i^2(m),
\end{aligned}$$

$$\begin{aligned}
P(1,2,0,0) &= \tau \sum_{q=1,Q} d(1,2,q) i^2(c), \\
P(1,1,1,1) &= 2\tau \sum_{q=1,Q} d(1,2,q) i(c)i(m), \\
P(1,0,2,2) &= \tau \sum_{q=1,Q} d(1,2,q) i^2(m).
\end{aligned}$$

The other three zero-order coefficients and the other six second-order coefficients are all zero:

$$\begin{aligned}
P(1,0,0,0) &= P(2,0,0,0) = P(3,0,0,0) = 0, \\
P(2,2,0,0) &= P(2,1,1,1) = P(2,0,2,0) = 0, \\
P(3,2,0,0) &= P(3,1,1,1) = P(3,0,2,0) = 0.
\end{aligned}$$

The amplitude and polarization coefficients are the real parts of the P coefficients

$$A(k,u,v,w) = \text{Re}(P(k,u,v,w)).$$

The retardation coefficients are the imaginary parts

$$\Phi(k,u,v,w) = \text{Im}(P(k,u,v,w)).$$

The polarization aberration coefficients are calculated for the foreoptics from the paraxial geometry and the normalized C vectors for the coatings and interfaces.

4.2.3 Polarization accuracy

The Jones vector gives the amplitude of the electric field and the square of the amplitude gives the intensity of the components. With the Jones matrix one is able to calculate the polarization effects for the optical system. The second-order Jones matrix which gives the linear polarization and linear retardance is given by

$$\begin{aligned}
JJ_2(H,\rho,\phi) &= \sigma(1)[H^2 P(1,2,0,0) + 2H\rho \cos(\phi) P(1,1,1,1) \\
&\quad + \rho^2 \cos(2\phi) P(1,0,2,2)] \\
&\quad + \sigma(2)[(2H\rho \sin(\phi) P(1,1,1,1) + \rho^2 \sin(2\phi) P(1,0,2,2))].
\end{aligned}$$

The on-axis linear polarization and linear retardance of the SAMEX foreoptics are characterized by the term linear defocus. The instrumental polarization function $JJ(H,\rho,\phi)$ for linear defocus is

$$\begin{aligned}
 JJ(H, \rho, \phi) &= \tau[\sigma_0 + P(1,0,2,2)\rho^2(\sigma_1 \cos 2\phi - \sigma_2 \sin 2\phi)] \\
 &= \tau[\sigma_0 + (a_{1022} + \tilde{\tau}\delta_{1022})\rho^2(\sigma_1 \cos 2\phi - \sigma_2 \sin 2\phi)].
 \end{aligned}$$

Here, τ is the amplitude transmittance of the system down the optical axis. It (τ) describes the polarization independent reflection and absorption losses associated with the ray down the optical axis at normal incidence at all interfaces. $P(1,0,2,2)$ describes the linear polarization (a_{1022}) and linear retardance (δ_{1022}) associated with the marginal ray.

Maximum coupling occurs when the incident light is circularly polarized, since circularly polarized light can always be decomposed into equal components of J_r and J_t everywhere in the pupil. The coupling is zero in the center of the pupil (where the polarization and retardance vanishes) and increases to a maximum coupling fraction of

$$I_{c,max}(H,1,\phi) = |P(1,0,2,2)|^2 = a_{1022}^2 + \delta_{1022}^2$$

at the edge of the pupil. The net fraction of incident circular polarized light coupled into the orthogonal circular polarized state is given by the integral over the pupil of

$$\begin{aligned}
 I_c &= \frac{|\tau|^2}{\pi} \int_0^{2\pi} d\phi \int_0^1 \rho d\rho |\rho^2 P(1,0,2,2)|^2 \\
 &= \frac{|\tau P(1,0,2,2)|^2}{3}.
 \end{aligned}$$

For incident linear or elliptical polarized light, the fraction of coupled intensity is less because the light is not composed of equal fractions of eigenstates.

The coupling is minimum for incident linear polarized light, which will be in one of the eigenpolarizations along one axis in the pupil and in the orthogonal eigenpolarization along the orthogonal axis. The fraction of coupled energy will be calculated assuming an incident polarization state of horizontal linear polarized light H for calculational simplicity, the same fraction is coupled for any linear polarized incident state. The polarization state transmitted by an optical system described by linear polarization defocus for incident H is

$$J(H, \rho, \phi) = \tau(\hat{H} + P(1,0,2,2) \rho^2(\hat{H} \cos 2\phi - \hat{V} \sin 2\phi)).$$

The fraction of incident \hat{H} light coupled into \hat{V} light is equal to

$$\begin{aligned}
 I_C &= \frac{|\tau|^2}{\pi} \int_0^{2\pi} d\phi \int_0^1 \rho d\rho |\vec{J}(H, \rho, \phi) \cdot \hat{V}|^2 \\
 &= \frac{|\tau|^2}{\pi} \int_0^{2\pi} d\phi \int_0^1 \rho d\rho |P(1, 0, 2, 2) \rho^2 \sin 2\phi|^2 \\
 &= \frac{|\tau P(1, 0, 2, 2)|^2}{\pi} \int_0^{2\pi} \sin^2 2\phi d\phi \int_0^1 \rho^5 d\rho \\
 &= \frac{|\tau P(1, 0, 2, 2)|^2}{6} .
 \end{aligned}$$

This is the minimum fraction of energy coupled by linear defocus aberrations. Since any elliptically polarized incident beam can be written as a sum of linear and circularly polarized light, the coupling fraction for arbitrarily polarized light lies in the range

$$\frac{|\tau P(1, 0, 2, 2)|^2}{6} < I_C < \frac{|\tau P(1, 0, 2, 2)|^2}{3} .$$

When unpolarized circularly polarized light is incident on the optical system, maximum polarization coupling occurs. This maximum is the polarization accuracy of a system and is calculated by the coupling integral, $I_{C, \max}$,

$$I_{C, \max} = \frac{1}{\pi} \int \int_{\text{pupil}} |c^2(1) + c^2(2) + c^2(3)| \rho d\rho d\phi .$$

This integral must be evaluated numerically except for some special cases (see section II.3.5). However, an analytic upper bound on the polarization accuracy can be easily established by using the triangular inequality

$$\begin{aligned}
 I_{C, \max} &< \frac{1}{\pi} \int \int_{\text{pupil}} [(\text{Re}[c(1)])^2 + (\text{Re}[c(2)])^2 + (\text{Re}[c(3)])^2]^2 \\
 &\quad + [(\text{Im}[c(1)])^2 + (\text{Im}[c(2)])^2 + (\text{Im}[c(3)])^2]^2 \rho d\rho d\phi .
 \end{aligned}$$

We can evaluate the upper bound on the polarization accuracy using the second-order polarization aberration expansion coefficients. The following integration is for the real part of the aberration coefficients, a_{1uvw} :

$$\begin{aligned}
I &= \frac{|\tau|^2}{\pi} \int_0^{2\pi} d\phi \int_0^1 \rho d\rho [(a_{1200}H^2 + a_{1111}H\rho \cos\phi + a_{1022}\rho^2 \cos 2\phi)^2 \\
&\quad + (a_{1111}H\rho \sin\phi + a_{1022}\rho^2 \sin 2\phi)^2] \\
&= |\tau|^2 [a_{1200}^2 H^4 + \frac{a_{1111}^2 H^2}{2} + \frac{a_{1022}^2}{3}].
\end{aligned}$$

The corresponding integral, $I_{\text{imaginary}}$, is obtained by substituting the imaginary part of the polarization aberration coefficients, δ_{1uvw} , for a_{1uvw} in the above expression. The upperbound on the polarization coupling is

$$\begin{aligned}
I_{c,\text{max}} &\leq I_{\text{real}} + I_{\text{imaginary}} \\
&= |\tau|^2 [a_{1200}^2 H^4 + \frac{a_{1111}^2 H^2}{2} + \frac{a_{1022}^2}{3} + \delta_{1200}^2 H^4 + \frac{\delta_{1111}^2 H^2}{2} + \frac{\delta_{1022}^2}{3}] \\
&= |\tau|^2 [|P(1,2,0,0)|^2 H^4 + \frac{|P(1,1,1,1)|^2 H^2}{2} + \frac{|P(1,0,2,2)|^2}{3}].
\end{aligned}$$

This is the second-order upper bound to the polarization coupling for systems of weak polarizers.

Hence the average effect over the image and pupil can be obtained by integrating over ρ and ϕ . The polarization accuracy, Δ , is defined as the maximum fraction of light (intensity)^p which can be coupled into an orthogonal polarization state. The incident polarized state is given by the Jones vector, \vec{J} . This vector is effectively rotated by the optical system and the rotation is given by the Jones matrix, JJ . The amount of polarization along the orthogonal state of polarization, \vec{J}' , of the incident polarization state is given by the projection of $JJ(\vec{J})$ into \vec{J}' ,

$$JJ(\vec{J}) \cdot \vec{J}'.$$

This value is given by the square of the second-order Jones matrix and is given in terms of the polarization aberration coefficients. (For the incident light in the polarization state I_p , then the result of instrumental polarization is to couple the polarization into the orthogonal state q . The amount of coupling is given by

$$I_{Op} \rightarrow I_O(1-\Delta_p)I_p + \Delta_p I_{Oq}$$

where Δ_p = polarization accuracy.) From the above results, the polarization accuracy is given in terms of the second-order polarization aberration coefficients $P(1,0,2,2)$, $P(1,1,1,1)$, and $P(1,2,0,0)$:

$$\Delta_p = \frac{1}{\pi} \int |JJ_2(H=1, \rho, \phi)|^2 \rho d\rho d\phi$$

or

$$\Delta_p = 2P^2(1,2,0,0) + \frac{1}{2}P^2(1,1,1,1) + \frac{1}{3}P^2(1,0,2,2)$$

where the integrations have been carried out for the squared terms and estimated for the crossed terms. For the SAMEX magnetograph design given herein the polarization accuracy value of

$$\Delta_p < 1.4 \times 10^{-7}$$

is obtained for specially selected optical coatings. These second-order aberration coefficients are given in section II.4. The second-order coefficients are sufficient since the next order that contributes is the fourth order. The polarization effects would be on the order of $(\Delta_p)^2$.

REFERENCES

- Azzam, R.M.A., and Bashara, N. M.: 1977, Ellipsometry and Polarized Light, North Holland, Amsterdam.
- Chipman, R. A.: 1987, "Polarization Aberrations," Ph.D. Dissertation, University of Arizona, Tucson, Arizona.
- Hecht, E., and Zajac, A.: 1974, Optics, Addison-Wesley, Reading, Massachusetts.
- Jones, R. C.: 1941a, J. Opt. Soc. Am. 31, 488.
- Jones, R. C.: 1941b, J. Opt. Soc. Am. 31, 493.
- Jones, R. C.: 1941c, J. Opt. Soc. Am. 31, 500.
- Jones, R. C.: 1942, J. Opt. Soc. Am. 32, 486.
- Jones, R. C.: 1947, J. Opt. Soc. Am. 37, 110.
- Shurcliff, W. A.: 1962, Polarized Light, Production and Use, Harvard University Press, Cambridge, MA.
- Theocaris, P. S., and Gdoutos, E. E.: 1979, Matrix Theory of Photoelasticity, Springer-Verlag, Berlin.

APPENDIX E

Interpretation of Data from the SAMEX Vector Magnetograph

1. Introduction

The SAMEX vector magnetograph is representative of those systems that use a narrow-band filter plus an assembly of polarizing optics to uniquely determine the state of polarization of a narrow wavelength interval of the light emerging from a small element of the Sun's surface. If we make the assumption that the polarization of the light is due solely to the presence of a homogeneous magnetic field throughout the height in the solar atmosphere where the absorption line is formed, then this magnetic field can be determined both in magnitude and direction from measurements of the degrees of linear and circular polarization of the light transmitted by the narrow-band filter and from the direction of the linear polarization. The minimum detectable line-of-sight and transverse magnetic fields are determined by the system's sensitivity for detection of fractional circular and linear polarizations, P_V and P_Q , respectively. The design goal for the SAMEX vector magnetograph is a polarimetric sensitivity of 10^{-4} .

The functional relationships between the fractional polarizations and magnetic field are derived using the solutions of Kjeldseth-Moe (1968) to the radiative transfer equations in the weak-field approximation. These relations take the form

$$\begin{aligned} B_L &= C_1 P_V, \\ \text{and} \\ B_T &= C_2 (P_Q)^{1/2}, \end{aligned}$$

where the constants C_1 and C_2 depend on the solar atmospheric model used in the Kjeldseth-Moe calculations. Examples of models used are the photospheric model of Holweger (1967), the plage model of Stenflo (1975), and the penumbral model of Kjeldseth-Moe and Maltby (1969). The constants C_1 and C_2 also depend on the line profile, the spectral filter's transmission characteristics, and the wavelength position of the filter relative to the spectral line. In this Appendix we will examine the assumptions implicit in using the equations given above to determine the line-of-sight and transverse components of the magnetic vector. To do this, we will first indicate how the fractional circular and linear polarizations P_V and P_Q are derived from the observational data. Then we will show how two of the components of the magnetic vector, the line-of-sight (B_L) and transverse (B_T) components, can be related to P_V and P_Q . The third component of the vector, the azimuth of the transverse field ϕ , is inferred from the plane of polarization of the linearly polarized light P_Q .

2. Polarization Analysis

The polarizing optics of the SAMEX magnetograph will consist of a removable quarter-wave plate and a rotating linear polarizer. The analysis of the incident polarized light for P_V , P_O , and ϕ can be accomplished with a minimum of five different configurations of the retarder and polarizer with this system. The effects of the quarter-wave plate and rotating polarizer on an incident beam of partially polarized light can be described in terms of the Stokes vector (I,Q,U,V) and the Mueller calculus (Shurcliff, 1962). The Stokes vector completely describes the intensity and polarization of a beam of polarized light. The first component of the vector, I, is the intensity of the light. The Q component represents the intensity of that portion of the light that is linearly polarized parallel to some fixed direction in the plane perpendicular to the direction of propagation of the light. The U component represents the intensity of that portion of the light that is linearly polarized at an angle of $+45^\circ$ to the fixed direction. Finally, V is the intensity of right-circularly polarized light. The Mueller calculus uses matrix algebra to describe the final state of polarization of a beam of light that has passed through a set of optical devices; the state of polarization is represented by a column matrix whose components are the components of the Stokes vector. Each element of the set of optical devices can be represented by a unique 4 by 4 matrix, and the matrices are arranged in the order of the optical elements with the matrix of the element last traversed by the light written at the left. An entire set of 'm' optical elements can thus be reduced mathematically to one 4 by 4 matrix which is the result of matrix multiplication of the m matrices. The process is shown schematically as follows:

$$\begin{pmatrix} I_i \\ Q_i \\ U_i \\ V_i \end{pmatrix} \xrightarrow{\text{optical element 1}} \xrightarrow{\text{optical element 2}} \dots \xrightarrow{\text{optical element m}} \begin{pmatrix} I \\ Q \\ U \\ V \end{pmatrix},$$

$$\begin{pmatrix} I \\ Q \\ U \\ V \end{pmatrix} = M_m \times M_{m-1} \times \dots \times M_2 \times M_1 \times \begin{pmatrix} I_i \\ Q_i \\ U_i \\ V_i \end{pmatrix}.$$

To use this formalism, the Stokes vector must be defined relative to a chosen coordinate system. We will choose the system defined by Unno (1956) where the z_0 axis is parallel to the direction of propagation of light and the magnetic vector lies in the x_0 - z_0 plane as shown in Figure E.1.

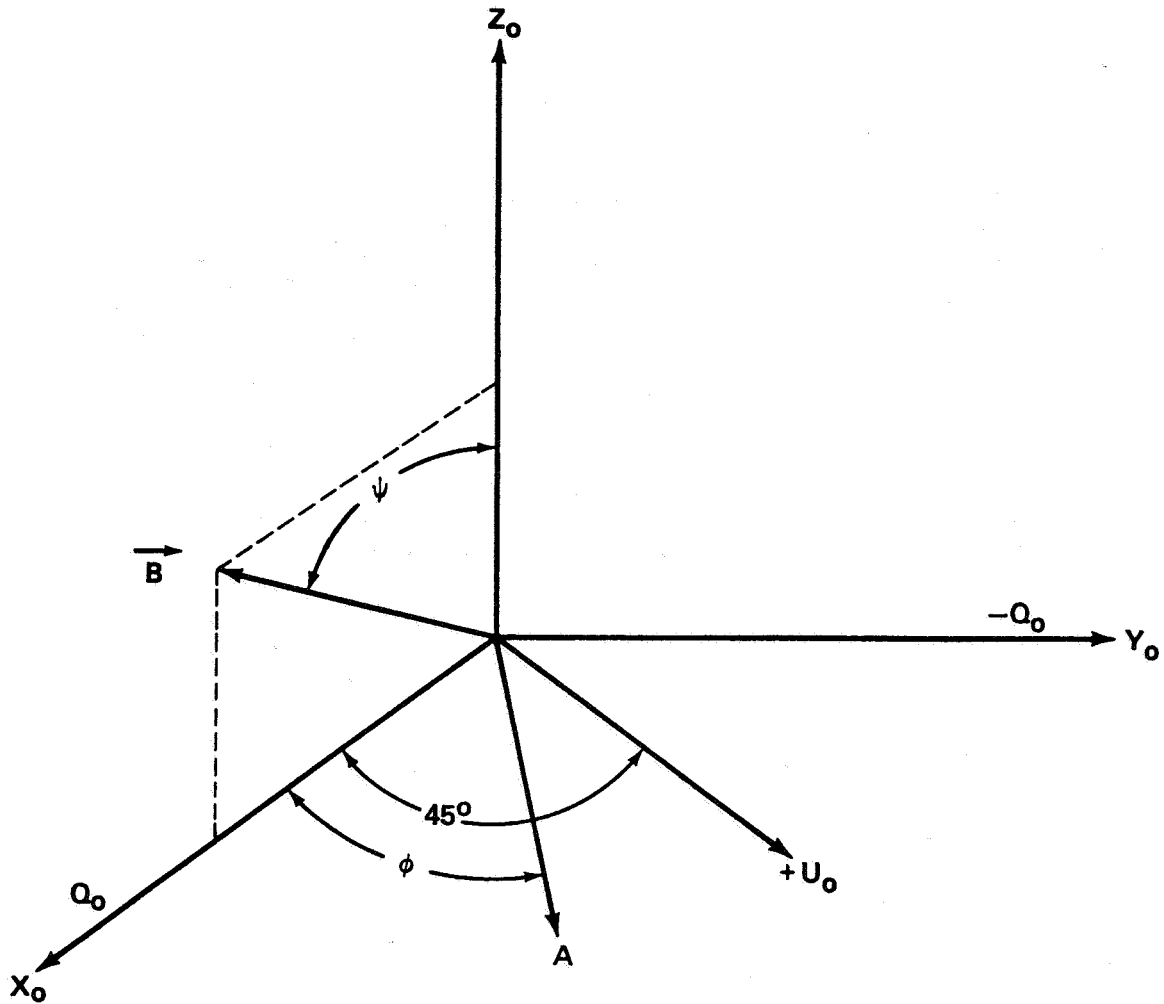


Figure E.1. The Unno coordinate system. The coordinate system x_0, y_0, z_0 in which Unno (1956) defined the Stokes vector is shown along with the two linear components Q_0 and U_0 of that vector. The magnetic field vector \vec{B} lies in the x_0-z_0 plane; its transverse component makes an angle ϕ with a designated direction A in the image plane.

The Mueller matrix for the SAMEX polarimeter is easily derived since there are only two elements, a quarter-wave plate with its fast axis fixed, and a rotating polarizer. The Mueller matrix for a linear retarder with retardance δ whose optical axis is oriented at an angle α is given by

$$M_{LR} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & D^2 - E^2 + G^2 & 2DE & -2EG \\ 0 & 2DE & -D^2 + E^2 + G^2 & 2DG \\ 0 & 2EG & -2DG & 2G^2 - 1 \end{pmatrix}.$$

In this matrix, the terms D, E, and G represent the following quantities:

$$D = \cos(2\alpha) \sin(\delta/2),$$

$$E = \sin(2\alpha) \sin(\delta/2),$$

$$G = \cos(\delta/2).$$

For light incident normally to the retarder's surface, the retardance δ is given by the expression

$$\delta = 2\pi \frac{d}{\lambda_o} (n_e - n_o)$$

where d is the thickness of the retarder, and n_e and n_o are the refractive indices of the crystal for the extraordinary and ordinary rays, respectively, at the wavelength λ_o . For off-axis rays that are incident at an angle i, where i is small, the retardance becomes

$$\delta \approx 2\pi \frac{d}{\lambda_o} (n_e - n_o) \left[1 - \frac{i^2}{2n_o} \left(\frac{\cos^2 w}{n_o} - \frac{\sin^2 w}{n_e} \right) \right],$$

where w is the angle that the plane of incidence makes with the optical axis of the retarder.

The Mueller matrix for a rotating polarizer whose transmittance axis makes an angle θ with a specified direction can be expressed as

$$M_{RP} = 1/2 \begin{pmatrix} 1 & C_2 & S_2 & 0 \\ C_2 & C_2^2 & C_2 S_2 & 0 \\ S_2 & C_2 S_2 & S_2^2 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

where $c_2 = \cos 2\theta$ and $S_2 = \sin 2\theta$.

Then, the Mueller matrix formulation for an optical system with a linear retarder followed by a rotating polarizer becomes

$$\begin{pmatrix} I \\ Q \\ U \\ V \end{pmatrix} = M_{RP} \times M_{LR} \times \begin{pmatrix} I_i \\ Q_i \\ U_i \\ V_i \end{pmatrix}.$$

When the linear retarder is removed from the optical path, only the matrix M_{RP} remains in the above equation.

In the design of the SAMEX polarimeter, the quarter-wave plate is aligned parallel to the rotating polarizer when the polarizer is set at an angle of 0° relative to some designated axis A; this axis will be the reference axis for all measurements of the azimuth of the solar vector magnetic field. Then, for a retardance of 90° and an optical axis aligned at 0° relative to the axis A, the terms in the matrix M_{LR} reduce to

$$E = 0,$$

$$D = G = \sqrt{2}/2.$$

In the first case, with the quarter-wave plate removed, the emergent intensity I is given by

$$I = 1/2 (I_i + Q_i \cos 2\theta + U_i \sin 2\theta),$$

where the incident linear Stokes parameters Q_i and U_i are defined with reference to the axis A.

In the second case, the quarter-wave plate is inserted into the optical path and the resulting intensity is

$$I = 1/2 (I_i + Q_i \cos 2\theta + V_i \sin 2\theta).$$

Rotation of the polarizer in steps of 45° then produces the emergent intensities that are given in Table 21 in section II.4 ("The Polarimeter"). Using the notation of that section, we have the following expressions for the five polarization measurements that are the minimum set of data required to determine the Stokes vector:

$$Sq_1 = I_i + Q_i ,$$

$$Sq_2 = I_i - Q_i ,$$

$$Su_1 = I_i + U_i ,$$

$$Sv_1 = I_i + V_i ,$$

$$Sv_2 = I_i - V_i .$$

In theory, the last measurement could be eliminated since the intensity derived from the first two measurements can be used in the third and fourth to derive U_i and V_i . However, in the actual operation of the magnetograph, we plan to enhance the circularly-polarized images a fewer number of times relative to the linear measurements, and so we need separate measurements of the intensity I_i for the circular and linear intensities.

We now want to express the polarization measurements Sx_i with reference to the Unno coordinate system. Referring to Figure E.1, the axis A is rotated an angle ϕ from the x_o axis of the coordinate system in which we will define the Stokes parameters I_o , Q_o , U_o , and V_o . But the transverse component of the magnetic vector lies along the x_o axis. Thus, ϕ is the angle of the axis A measured relative to the transverse component of the magnetic field that produces the polarized intensities Q_o , U_o , and V_o ; in other words, ϕ is the azimuth of the magnetic field and one of the quantities we wish to measure. Relative to the Unno coordinate system, then, the angles α and θ in the matrices M_{LR} and M_{RP} become ϕ and $\phi + \theta$, respectively, with ϕ measured with respect to x_o and θ with respect to A. Thus, the terms in the matrix M_{LR} now become

$$D = (\sqrt{2}/2) \cos 2\phi ,$$

$$E = (\sqrt{2}/2) \sin 2\phi ,$$

$$G = \sqrt{2}/2 .$$

The intensity emerging from the SAMEX polarimeter when the quarter-wave plate is out of the optics is now given by the expression

$$I = 1/2 [I_o + Q_o \cos 2(\phi + \theta) + U_o \sin 2(\phi + \theta)] .$$

With the wave plate in the optical path, the intensity becomes

$$\begin{aligned}
I = & 1/2 \{ I_0 + Q_0 [\cos^2 2\phi \cos 2(\phi + \theta) + \sin 2\phi \cos 2\phi \sin 2(\phi + \theta)] \\
& + U_0 [\sin^2 2\phi \sin 2(\phi + \theta) + \sin 2\phi \cos 2\phi \cos 2(\phi + \theta)] \\
& + V_0 [\cos 2\phi \sin 2(\phi + \theta) - \sin 2\phi \cos 2(\phi + \theta)] \} .
\end{aligned}$$

The five needed measurements are summarized in the following table.

Table E.1 Measured polarizations as functions of the Stokes parameters.

Polarizer Angle	Polarization Measurement	$\lambda/4$ Plate	Stokes Component
0°	Sq ₁	out	1/2 (I ₀ + Q ₀ cos 2φ + U ₀ sin 2φ)
90°	Sq ₂	out	1/2 (I ₀ - Q ₀ cos 2φ - U ₀ sin 2φ)
45°	Su ₁	out	1/2 (I ₀ - Q ₀ sin 2φ + U ₀ cos 2φ)
45°	Sv ₁	in	1/2 (I ₀ + V ₀)
135°	Sv ₂	in	1/2 (I ₀ - V ₀)

Following the polarizing optics, the light path is directed through the spectral filter centered at a selected wavelength offset λ_i from the central wavelength of a Zeeman-sensitive absorption line. The signal $\overline{Sx_i}$ that is finally incident on the SAMEX detector is thus the intensity Sx_i integrated over the narrow bandpass 2Δ of the filter whose transmission profile is $T(\lambda_i, \lambda)$:

$$\overline{Sx_i} = \int_{\lambda_i - \Delta}^{\lambda_i + \Delta} Sx_i(\lambda) T(\lambda, \lambda_i) d\lambda .$$

In the theoretical treatment of the radiative transfer of the polarized light, we assume that the magnetic field is homogeneous throughout the line-forming region of the solar atmosphere so that the azimuth ϕ of the magnetic field does not change with height. Hence, if we neglect Faraday rotation, the plane of linear polarization of the light transmitted through the line-forming region is also unchanged. But this plane is

created by the magnetic field, so that U_0 can be set equal to zero in the expressions for the measured intensities, and the net linear polarization is contained entirely in the Q_0 component.

With this approximation, we now derive the three quantities that we will relate to the three components of the vector magnetic field, namely, the fractional linear polarization P_Q , the fractional circular polarization P_V , and the azimuth of the field ϕ . These three quantities are derived from the filtered intensities Sx_i as follows:

$$\begin{aligned} P_Q &= (Sq_x^2 + Sq_y^2)^{1/2}, \\ P_V &= Sv, \\ \phi &= 1/2 \tan^{-1} (Sq_y/Sq_x), \end{aligned} \quad (1)$$

where

$$\begin{aligned} Sq_x &= (\overline{Sq_1} - \overline{Sq_2})/(\overline{Sq_1} + \overline{Sq_2}) = (\overline{Q_0}/\overline{I_0}) \cos 2\phi, \\ Sq_y &= (\overline{Sq_1} + \overline{Sq_2} - 2 \overline{Su_1})/(\overline{Sq_1} + \overline{Sq_2}) = (\overline{Q_0}/\overline{I_0}) \sin 2\phi, \end{aligned}$$

and

$$Sv = (\overline{Sv_1} - \overline{Sv_2})/(\overline{Sv_1} + \overline{Sv_2}) = \overline{V_0}/\overline{I_0}. \quad (2)$$

We now must derive the relation between the line-of-sight and transverse components of the magnetic field and the two quantities, P_V and P_Q . The third component of the field, the azimuth, is equal to ϕ .

3. Radiative Transfer Equations of the Stokes Parameters

The basic radiative transfer equation without polarization effects for a ray of radiation passing through an elemental volume on the Sun is given by (Mihalas, 1970):

$$\mu \frac{dI_\lambda}{d\tau} = I_\lambda - S_\lambda.$$

In this equation, $\mu = \cos \theta$ with θ equal to the zenith angle, S_λ is the source function (the ratio of emissivity to opacity), and the differential optical depth $d\tau$ is equal to

$-K_{\lambda}^C dz$, where K_{λ}^C is the continuum absorption coefficient or opacity and z is the vertical coordinate. This transfer equation simply states that the change in intensity of the radiation is the difference of the incident and exit radiation from the elemental volume, taking into account all absorption and emission processes that take place as the radiation passes through the volume. In the presence of a magnetic field, the inverse Zeeman effect complicates the radiative transfer processes by altering the exit radiation characteristics, depending on the strength and direction of the magnetic field.

The general differential equations for the radiative transfer of the Stokes intensities were derived (classically) by Unno (1956) for the case of a homogeneous magnetic field and the line formation mechanism of pure absorption. Kjeldseth-Moe (1968) extended Unno's theory to include scattering, and Beckers (1969) relaxed the assumption of a homogeneous magnetic field. Beckers (1969) also included the effects of anomalous dispersion that lead to Faraday rotation effects. Landi Degl'Innocenti (1976) generalized the transfer equations for pure absorption to include departures from LTE. These latter equations are the following:

$$\mu \frac{dI}{d\tau} = (I - B_T) + \eta_I(I - S) + \eta_Q Q + \eta_U U + \eta_V V ,$$

$$\mu \frac{dQ}{d\tau} = \eta_Q(I - S) + (1 + \eta_I) Q + \rho_V U - \rho_U V ,$$

$$\mu \frac{dU}{d\tau} = \eta_U(I - S) + (1 + \eta_I) U - \rho_V Q + \rho_Q V ,$$

$$\mu \frac{dV}{d\tau} = \eta_V(I - S) + (1 + \eta_I) V + \rho_U Q - \rho_Q U .$$

Here, S is the line source function (without scattering); for LTE, $S = B_T$, the Planck function. The coefficients η describe the relative contributions of the Zeeman components to the Stokes intensities, while the coefficients ρ describe the effects of anomalous dispersion. These coefficients are defined as follows:

$$\eta_I = \frac{\eta_p}{2} \sin^2 \psi + \frac{\eta_r + \eta_\ell}{4} (1 + \cos^2 \psi) ,$$

$$\eta_Q = \left(\frac{\eta_p}{2} - \frac{\eta_r + \eta_\ell}{4} \right) \sin^2 \psi \cos 2\phi_0 ,$$

$$\eta_U = \left(\frac{\eta_p}{2} - \frac{\eta_r + \eta_\ell}{4} \right) \sin^2 \psi \sin 2\phi_0 ,$$

$$\eta_V = \frac{\eta_r - \eta_\ell}{2} \cos \psi ,$$

$$\rho_Q = \left(\frac{\rho_p}{2} - \frac{\rho_r + \rho_\ell}{4} \right) \sin^2 \Psi \cos 2\phi_0 ,$$

$$\rho_U = \left(\frac{\rho_p}{2} - \frac{\rho_r + \rho_\ell}{4} \right) \sin^2 \Psi \sin 2\phi_0 ,$$

$$\rho_V = \frac{\rho_r - \rho_\ell}{2} \cos \Psi .$$

In these equations, Ψ is the inclination of the field to the line-of-sight (see Figure E.1), and ϕ_0 is the angle the field vector makes with the X_0 axis in Figure E.1 when projected onto the x_0 - y_0 plane. In the Unno coordinates, $\phi_0 = 0$, but ϕ_0 can vary with depth if the field B is not homogeneous. The functions $\eta_{p,r,\ell}$ and $\rho_{p,r,\ell}$ are defined by the following expressions:

$$\eta_p = \frac{K_\lambda^\ell(0)}{K_\lambda^c} H(a, \lambda) = \eta_0 H(a, \lambda) ,$$

$$\eta_{\ell,r} = \eta_0 H(a, \lambda \pm \Delta\lambda_B) ,$$

$$\rho_p = 2\eta_0 F(a, \lambda) ,$$

$$\rho_{\ell,r} = 2\eta_0 F(a, \lambda \pm \Delta\lambda_B) .$$

In these expressions, H and F are the Voigt and Faraday-Voigt functions with damping a , $K_\lambda^\ell(0)$ is the line absorption coefficient at line center (λ_0), and $\Delta\lambda_B$ is the Zeeman splitting factor. If we assume that the absorption line is broadened by a combination of natural, collisional (a) and Doppler ($\Delta\lambda_D$) broadening, the line absorption coefficient at line center, corrected for stimulated emissions, can be written as

$$K_\lambda^\ell(0) = \frac{\sqrt{\pi} e^2}{m_e c^2} \frac{\lambda_0^2 f}{\Delta\lambda_D} N_1 (1 - e^{-hc/\lambda_0 kT}) .$$

In this equation, f is the oscillator strength of the transition, the Doppler width is $\Delta\lambda_D = \frac{\lambda_0}{c} \left(\frac{2kT}{M} + \xi^2 \right)^{1/2}$ (M = atomic mass, ξ = microturbulent velocity field in the solar atmosphere), and N_1 is the number density of atoms which are in the absorbing state.

3.1 The Unno Solutions to the Transfer Equations

In his solutions to the differential equations for the Stokes parameters, Unno (1956) assumes (1) a homogeneous magnetic field so that the angle ϕ_0 is always zero; (2) a Milne-Eddington atmosphere so that if B is homogeneous $\eta_{I, O, U, V}$ are independent of τ ; (3) the line is formed through the mechanism of true absorption in LTE so that $S = B_T$, where B_T is the Planck intensity distribution

$$B_T = \frac{2hc^2}{\lambda_0^5} (e^{hc/\lambda_0 kT} - 1)^{-1} ;$$

(4) $B_T = B_0(1 + \beta_0 \tau)$; (5) anomalous dispersion is neglected; and (6) the boundary conditions are

$$\begin{aligned} Q &= U = V \rightarrow 0 \\ I &\rightarrow \text{continuum} \end{aligned} \quad \text{as } \tau \rightarrow \infty.$$

The fundamental solutions obtained by Unno for the emergent intensities ($\tau = 0$) are

$$I_O(0, \mu) = B_0 \left[1 + \mu \beta_0 \frac{(1 + \eta_I)}{(1 + \eta_I)^2 - D^2} \right] ,$$

$$Q_O(0, \mu) = -B_0 \mu \beta_0 \frac{\eta_Q}{(1 + \eta_I)^2 - D^2} ,$$

$$U_O(0, \mu) = 0 ,$$

$$V_O(0, \mu) = -B_0 \mu \beta_0 \frac{\eta_V}{(1 + \eta_I)^2 - D^2} ,$$

$$I_C(0, \mu) = \text{continuum intensity} = B_0(1 + \mu \beta_0),$$

and

$$D^2 = \eta_Q^2 + \eta_V^2 .$$

We are now ready to relate these emergent intensities to the two quantities P_V and P_O derived from the filtered intensities $\overline{Sx_i}$. From Equations (1) and (2), we have

$$P_O = \overline{Q_O} / \overline{I_O}$$

and

$$P_V = \overline{V_O} / \overline{I_O} ,$$

where the symbol $\overline{\quad}$ denotes that the intensity is integrated over the passband of the filter. Changing to the dimensionless variable $v = \Delta\lambda/\Delta\lambda_D$, we can write

$$P_V = \frac{-\mu\beta_O \int_{v_i-\delta}^{v_i+\delta} \frac{\eta_V(v)}{[1 + \eta_I(v)]^2 - D^2(v)} \cdot T(v, v_i) dv}{\int_{v_i-\delta}^{v_i+\delta} \left\{ 1 + \mu\beta_O \frac{(1 + \eta_I)}{[1 + \eta_I(v)]^2 - D^2(v)} \right\} \cdot T(v, v_i) dv} , \quad (3)$$

$$P_O = \frac{-\mu\beta_O \int_{v_i-\delta}^{v_i+\delta} \frac{\eta_Q(v)}{(1 + \eta_I)^2 - D^2(v)} \cdot T(v, v_i) dv}{\int_{v_i-\delta}^{v_i+\delta} \left\{ 1 + \mu\beta_O \frac{[1 + \eta_I(v)]}{[1 + \eta_I(v)]^2 - D^2(v)} \right\} \cdot T(v, v_i) dv} , \quad (4)$$

with $T(v, v_i)$ denoting the transmission profile of the spectral filter, and $\delta = \Delta/\Delta\lambda_D$. Thus, the measured quantities P_V , P_O can be related (indirectly) to the field strength B and inclination ψ , assuming that we have appropriate values for the Fraunhofer line parameters η_O , a , $\Delta\lambda_D$, and β_O and that we know the transmission profile $T(v, v_i)$ of the spectral filter.

3.2 Weak-Field Approximation

Equation (3) and (4) relating B to the measured quantities $P_{V,O}$ can be simplified for quick analysis of magnetograph data. We assume that v_B is small ($v_B \ll 1$) and expand the Voigt functions about the zero-field wavelength displacement v :

$$\eta_p = \eta_o H(a, v) ,$$

$$\eta_l = \eta_o H(a, v + v_B) = \eta_o [H(a, v) + v_B \frac{dH(a, v)}{dv} \Big|_v + \frac{1}{2} v_B^2 \frac{d^2 H(a, v)}{dv^2} \Big|_v - \dots] ,$$

or

$$\eta_l = \eta_p + \eta_o [v_B \frac{dH(a, v)}{dv} \Big|_v + \frac{1}{2} v_B^2 \frac{d^2 H(a, v)}{dv^2} \Big|_v + \dots] ,$$

$$\eta_r = \eta_o H(a, v - v_B) = \eta_o [H(a, v) - v_B \frac{dH(a, v)}{dv} \Big|_v + \frac{1}{2} v_B^2 \frac{d^2 H(a, v)}{dv^2} \Big|_v + \dots] ,$$

or

$$\eta_r = \eta_p + \eta_o [-v_B \frac{dH(a, v)}{dv} \Big|_v + \frac{1}{2} v_B^2 \frac{d^2 H(a, v)}{dv^2} \Big|_v + \dots] .$$

The coefficients $\eta_{I, Q, V}$ then become:

$$\eta_I = \eta_o H(a, v) + \left(\frac{1 + \cos^2 \psi}{4} \right) \left[\eta_o v_B^2 \frac{d^2 H}{dv^2} + 0 (v_B^4) \right] ,$$

$$\eta_Q = - \left[\eta_o v_B^2 \frac{d^2 H}{dv^2} + 0 (v_B^4) \right] \left(\frac{\sin^2 \psi}{4} \right) ,$$

$$\eta_V = \left[- \eta_o v_B \frac{dH}{dv} + 0 (v_B^3) \right] (\cos \psi) ,$$

$$D^2 = \left(\eta_o v_B \frac{dH}{dv} \cos \psi \right)^2 + 0 (v_B^4) .$$

Then, to the first order of v_B , equation (3) reduces to

$$P_V = (v_B \cos \psi) C_1' ,$$

where

$$C_1' = \frac{+\mu\beta_o \eta_o \int_{v_i-\delta}^{v_i+\delta} \frac{[\frac{dH(a,v)}{dv}]_v}{[1+\eta_o H(a,v)]^2} T(v, v_i) dv}{\int_{v_i-\delta}^{v_i+\delta} [1 + \frac{\mu\beta_o}{1+\eta_o H(a,v)}] T(v, v_i) dv}.$$

To approximate P_Q we must retain dominant terms of the order v_B^2 ; then

$$P_Q = (v_B \sin \psi)^2 C_2'$$

where

$$C_2' = \frac{\frac{1}{4}\mu\beta_o \eta_o \int_{v_i-\delta}^{v_i+\delta} \frac{[\frac{d^2H(a,v)}{dv^2}]_v}{[1+\eta_o H(a,v)]^2} T(v, v_i) dv}{\int_{v_i-\delta}^{v_i+\delta} [1 + \frac{\mu\beta_o}{1+\eta_o H(a,v)}] T(v, v_i) dv}.$$

The "constants" C_1' and C_2' are dependent on the line profile parameters ($\eta_o, a, \beta_o, \Delta\lambda_D$) and the filter's transmission profile and "setting" (v_i); the constants C_1' and C_2' can thus be calculated if these parameters can be determined independently. One difficulty which arises in connection with the line profile parameters is the change in the line profile in various active regions of the Sun as reported by Harvey and Livingston (1969). They have found that the Fe 5250.22 Å line, for example, weakens greatly in active regions (magnetic regions) outside sunspots. We must thus change the line profile parameters as we go from the "quiet" photosphere into active regions outside of sunspots as well as into penumbral and umbral regions. Thus, the "constants" C_1' and C_2' will depend on the type of magnetic region we are observing.

With these factors in mind we can now relate magnetograph signals (P_Q, P_V) to the Sun's magnetic field strength B and inclination ψ :

$$v_B \cos \psi = P_V / C_1', \quad (5)$$

$$v_B \sin \psi = (P_Q / C_2')^{1/2}. \quad (6)$$

We can write for v_B (using Giorgi M.K.S. units)

$$v_B = \frac{1}{\Delta\lambda_D} \left(\frac{e\lambda_o^2 g}{4\pi m_e c} \right) B ,$$

where g is the Landé g factor for the transition. Values of the term $(g\lambda^2)$ for various suitable Fraunhofer lines can be found in Bray and Loughhead (1964). If $(g\lambda^2)$ is in centimeters², $\Delta\lambda_D$ in meters (or milli-Angstrom), and we express B in teslas (or gauss), then

$$v_B = (4.67 \times 10^6) \left(\frac{g\lambda^2}{\Delta\lambda_D} \right) B .$$

For Fe 5250.22 Å

$$g\lambda^2 = 83 \times 10^{-10} \text{ cm}^2$$

which gives

$$v_B = \frac{38.76 \times 10^{-3}}{\Delta\lambda_D} B .$$

If $\Delta\lambda_D \approx 38.7 \times 10^{-6} \text{ m}$ (38.7 mÅ), then

$$v_B \ll 1 \text{ for } B \ll 1000 \text{ G.}$$

Substituting for v_B in Equations (5) and (6), we have finally

$$B_L = B \cos\psi = P_V \left(\frac{C_0'}{C_1} \right) \equiv C_1 P_V , \quad (7)$$

$$B_T = B \sin\psi = \frac{C_0'}{\sqrt{C_2}} P_Q^{1/2} = C_2 P_Q^{1/2} , \quad (8)$$

where

$$C_0' = \frac{\Delta\lambda_D}{4.67 \times 10^6 (g\lambda_o^2)} ,$$

and we have introduced the unprimed coefficients C_1 and C_2 . Equations (7) and (8) thus allow us to easily relate measured quantities to the longitudinal and transverse components of the Sun's magnetic field, provided that $v_B \ll 1$.

3.3 Evaluation of the Constants C_1' , C_2' , C_0'

The parameters η_0 , β_0 , and $\Delta\lambda_D$ which appear explicitly in the defining equations for the constants C_1' , C_2' , and C_0' , as well as the damping parameter a in the Voigt function, can be determined fairly well by fitting the theoretical line profile for zero magnetic field to an experimental profile taken in the "quiet" Sun. From the solutions for the emergent intensities I_0 and I_c , the case of zero magnetic field gives

$$\frac{I_0(v)}{I_c} = \frac{[1 + \frac{\mu\beta_0}{1+\eta_0 H(a,v)}]}{(1+\mu\beta_0)} .$$

The parameters in this equation are then varied until a good fit to the experimental profile is obtained.

Values obtained in this manner for the parameters η_0 , β_0 , $\Delta\lambda_D$ should be consistent with independent estimates of their magnitude. For example, the Doppler width $\Delta\lambda_D$ should be of the same order of magnitude as

$$\frac{\lambda_0}{c} \left(\frac{2kT}{M} + \xi^2 \right)^{1/2}$$

for representative values of T and ξ . Also, the parameter β_0 should be consistent with data on center-to-limb variations of the intensity of the line and continuum. The analytical expression for η_0 comes from the definition

$$\eta_0 = \frac{K_\lambda^l(0)}{K_\lambda^c},$$

so that

$$\eta_0 = \frac{\sqrt{\pi} e^2}{m_e c^2} \frac{\lambda_0^2 f}{\Delta\lambda_D} \frac{N_1}{K_\lambda^c} (1 - e^{-hc/\lambda_0 kT}) .$$

Assuming local thermodynamic equilibrium, the population density of the lower level, N_1 , can be obtained from the Saha and Boltzmann equations as a function of temperature T and electronic pressure P_e . Thus, using values of T , P_e , and K_λ^c representative of the line-forming depths of the solar atmosphere, the approximate range of values of η_0 can be calculated. Finally, we can make theoretical estimates of the damping factor a . Postma (private communication, 1970) has made such a calculation assuming van der Waals broadening for Fe 5250.22 Å, with the result that $\gamma/\gamma_D \approx 0.02$; using the usual enhancement factor of about 4, the value for a is ≈ 0.08 .

The derivatives of the Voigt function which occur in the equations for C_1' and C_2' can be evaluated by using the following approximate form for the Voigt profile (Avery and House, 1968):

$$|v| < 1.00 : H(a,v) \approx \frac{\pi}{\pi+2a} e^{-v^2} ,$$

$$|v| > 1.00 : H(a,v) \approx \frac{\pi}{\pi+2a} (e^{-v^2} + \frac{a}{\sqrt{\pi}v^2}) .$$

From these we have:

$$|v| < 1.00 : \frac{dH}{dv} \approx \frac{-2\pi}{\pi+2a} (ve^{-v^2}) ;$$

$$\frac{d^2H}{dv^2} \approx \frac{-2\pi}{\pi+2a} [(1-2v^2) e^{-v^2}] ;$$

$$|v| > 1.00 : \frac{dH}{dv} \approx \frac{-2\pi}{\pi+2a} (ve^{-v^2} + \frac{a}{\sqrt{\pi}v^3}) ;$$

$$\frac{d^2H}{dv^2} \approx \frac{-2\pi}{\pi+2a} [(1-2v^2) e^{-v^2} - \frac{3a}{\sqrt{\pi}v^4}] .$$

3.4 The Kjeldseth-Moe Solutions to the Transfer Equations

Kjeldseth-Moe (1968) obtained integral representations for the solutions to the transfer equations under the assumptions that (1) the magnetic field is homogeneous; (2) $H(a,v)$ is independent of depth; and (3) $Q = U = V \rightarrow 0$ and $I \rightarrow I_c$ as $\tau \rightarrow \infty$. If we add the further restriction that the line is formed by pure absorption, his solutions assume the following simple form:

$$I_o = \frac{(Z_1 + Z_2)}{2} ,$$

$$Q_o, V_o = \frac{\eta_{Q,V}}{D} \frac{(Z_1 - Z_2)}{2} ,$$

where

$$Z_{1,2} = \int_0^{\infty} (1 + \eta_I \pm D) B_T(\tau) e^{-\int_0^{\tau} (1 + \eta_I \pm D) \frac{d\tau'}{\mu}} \frac{d\tau}{\mu} .$$

The continuum intensity is found by setting $\eta_I = D = 0$ in this equation. A detailed description of the numerical evaluation of the integrals involved in the equations for Z_1 and Z_2 has been given by Hagyard (1970). The numerical solutions found for I_o , Q_o , V_o , and I_c are then integrated over the filter passband to calculate P_V and P_o . The undetermined parameters involved in this analysis are $\Delta\lambda_D$, ξ , a , and fA_e , where A_e is the abundance of the element; the choice of model atmosphere is also somewhat arbitrary.

The application of these solutions to magnetograph calibration is not as straightforward as in the case of the Unno analysis for $v_B \ll 1$. However, "calibration curves" can be obtained by calculating P_o and P_V for various values of B and ψ and plotting the results of P_o versus B and P_V versus B for values of ψ running from 0° to 90° . For small ($\lesssim 1000$ G) magnetic fields, these curves can be used to determine a unique B and ψ from measurements of P_o and P_V ; for larger field strengths the interpretations are somewhat ambiguous because the calibration curves become multi-valued.

3.5 More General Analytical Solutions for a Milne-Eddington Atmosphere

In the solutions proposed by Unno, the assumption that the source function B_T varies linearly with depth is known to be inaccurate because it requires that the center-limb variation of the continuum intensity follow the relation $I_c(\mu) = B_o(1 + \beta_o\mu)$. Observational data show this to be an inaccurate description, especially near the limb. Better approximations to the variation of B_T with depth are (Pecker, 1965):

$$B_T(\tau) = a + b\tau + c\tau^2 \dots ,$$

$$B_T(\tau) = a + b\tau + ce^{-k\tau} ,$$

$$B_T(\tau) = a + b\tau + cE_2(\tau) + dE_3(\tau) .$$

Hagyard (1971) has shown how the Kjeldseth-Moe solutions can be adapted to the more restricted problem of a Milne-Eddington atmosphere to yield analytical solutions for the Stokes parameters I_0 , Q_0 , V_0 , and I_c for non-linear source functions like these. Using these results we can construct calibration curves without resorting to a model atmosphere.

3.6 Theoretical Calibration Curves

Using these various atmospheric models, computed calibration curves give the variations of P_V and P_Q with field strength B as functions of atmospheric model, inclination Ψ , disk position μ , and the wavelength of peak filter transmission relative to the spectral line center; Figures E.2 and E.3 show examples of these curves. From these, the region of validity for the weak field approximations can be determined as well as the constants C_1 and C_2 .

For strong magnetic fields, the interpretation of magnetograph data becomes more difficult because of the multi-valued nature of the calibration curves. For these fields, data from observations at different filter spectral positions must be examined and interpreted using calibration curves derived for variable filter positions as shown in Figure E.4. Examination of curves such as these indicates that the region of validity for the linear relations (Equations (7) and (8)) can be extended to higher field strengths by making observations further into the wings of the spectral line. However, the constants C_1 and C_2 are functions of this filter spectral position. In Table E.2 these constants are given for a penumbral model at different filter positions; the respective ranges of validity are also indicated in gauss, to an accuracy of ± 100 gauss for the higher field strengths.

In interpreting data from a magnetogram covering an active region, the variations in the parameters C_1, C_2 from photosphere to penumbra and umbra must be considered in deriving relative field intensities (effects of line weakening in areas underlying H-alpha plage regions have also been demonstrated by Hagyard and Teuber, 1978). However, with the exception of umbral models, the ratio C_2/C_1 is approximately the same for a given filter position throughout an active region area as indicated in Table E.3. Thus, from Equations (7) and (8), the inclination Ψ , given by

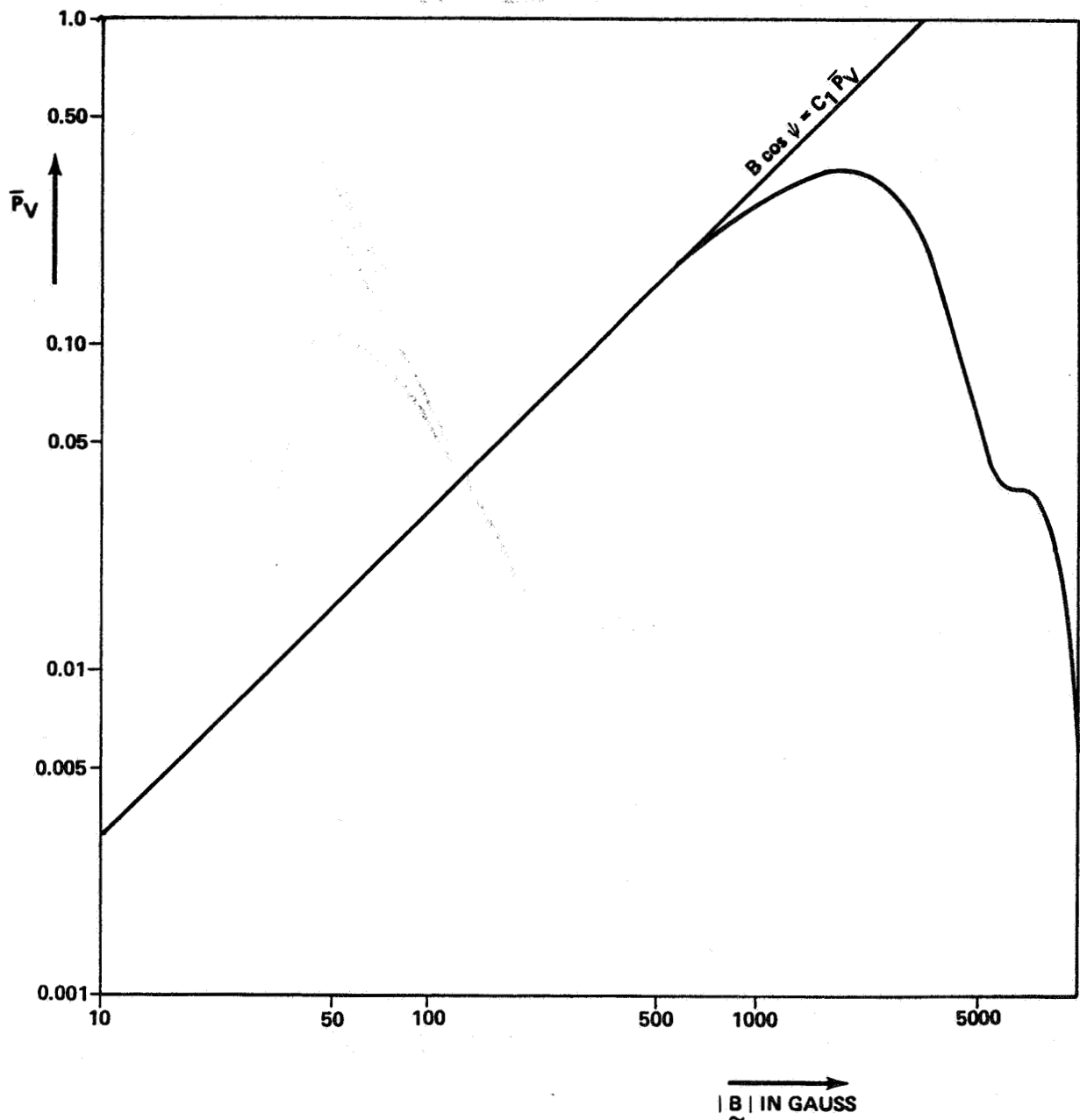


Figure E.2. Calibration curve relating measured circular polarization to the magnitude of the magnetic field. The calculated fractional circular polarization P_V is plotted as a function of field strength for a field that lies along the line of sight ($\psi = 0^\circ$). The analysis of Kjeldseth-Moe (1968) was used with a penumbral model for the solar atmosphere; the circular polarization is the integrated signal transmitted by a $1/8 \text{ \AA}$ bandpass birefringent filter centered at 60 m\AA in the blue wing of the Fe I 5250 \AA spectral line. The straight line represents the linear approximation relating the line-of-sight component B_L to circular polarization.

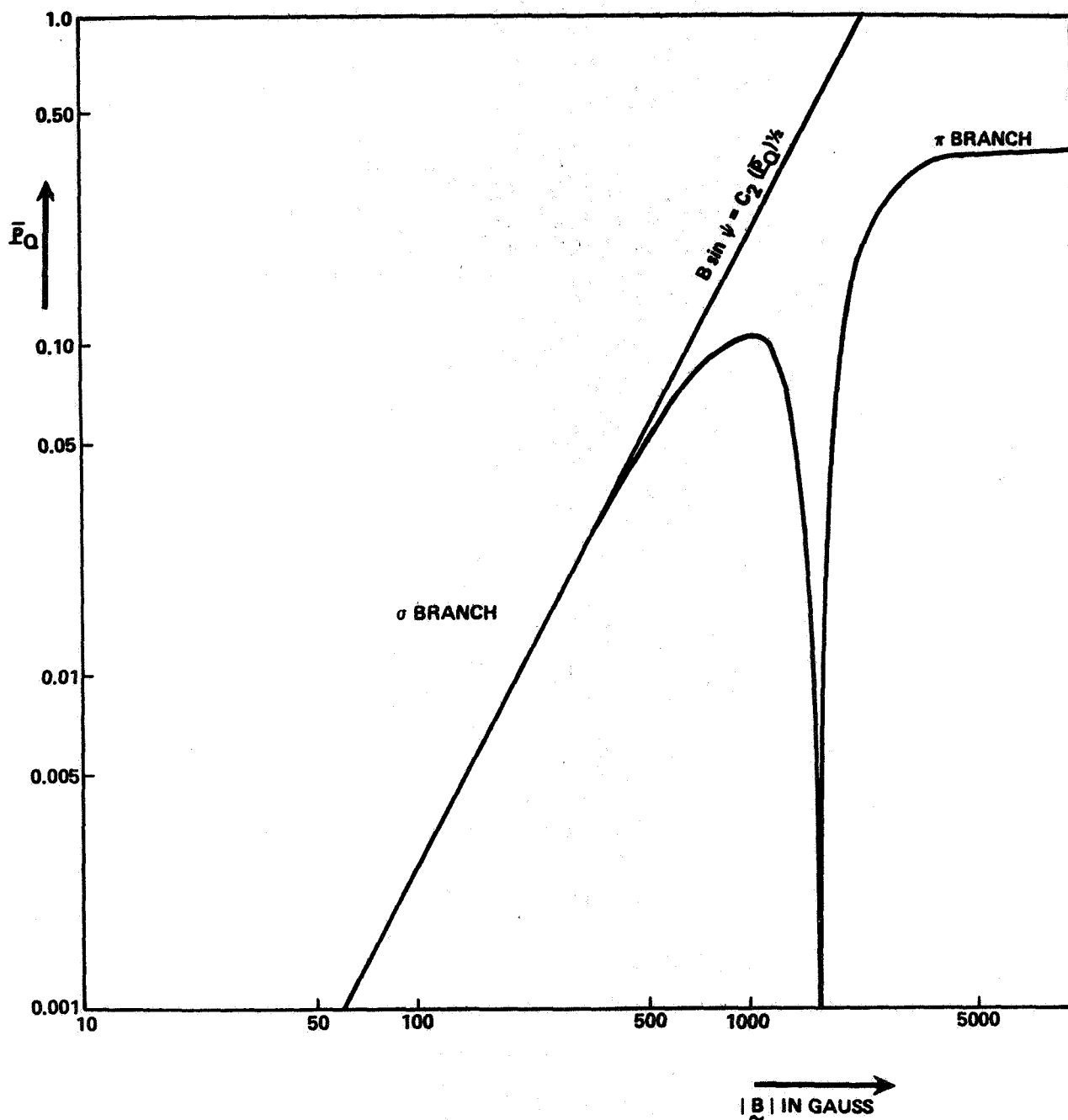


Figure E.3. Calibration curve relating measured linear polarization to the magnitude of the magnetic field. The calculated fractional linear polarization P_0 is plotted as a function of field strength for a field that lies transverse to the line of sight ($\psi = 90^\circ$). The analysis of Kjeldseth-Moe (1968) was used with a penumbral model; the linear polarization is the integrated signal transmitted by a $1/8 \text{ \AA}$ bandpass birefringent filter centered at the center of the Fe I 5250 \AA spectral line. The straight line represents the quadratic approximation relating the transverse component B_T to the square root of the linear signal.

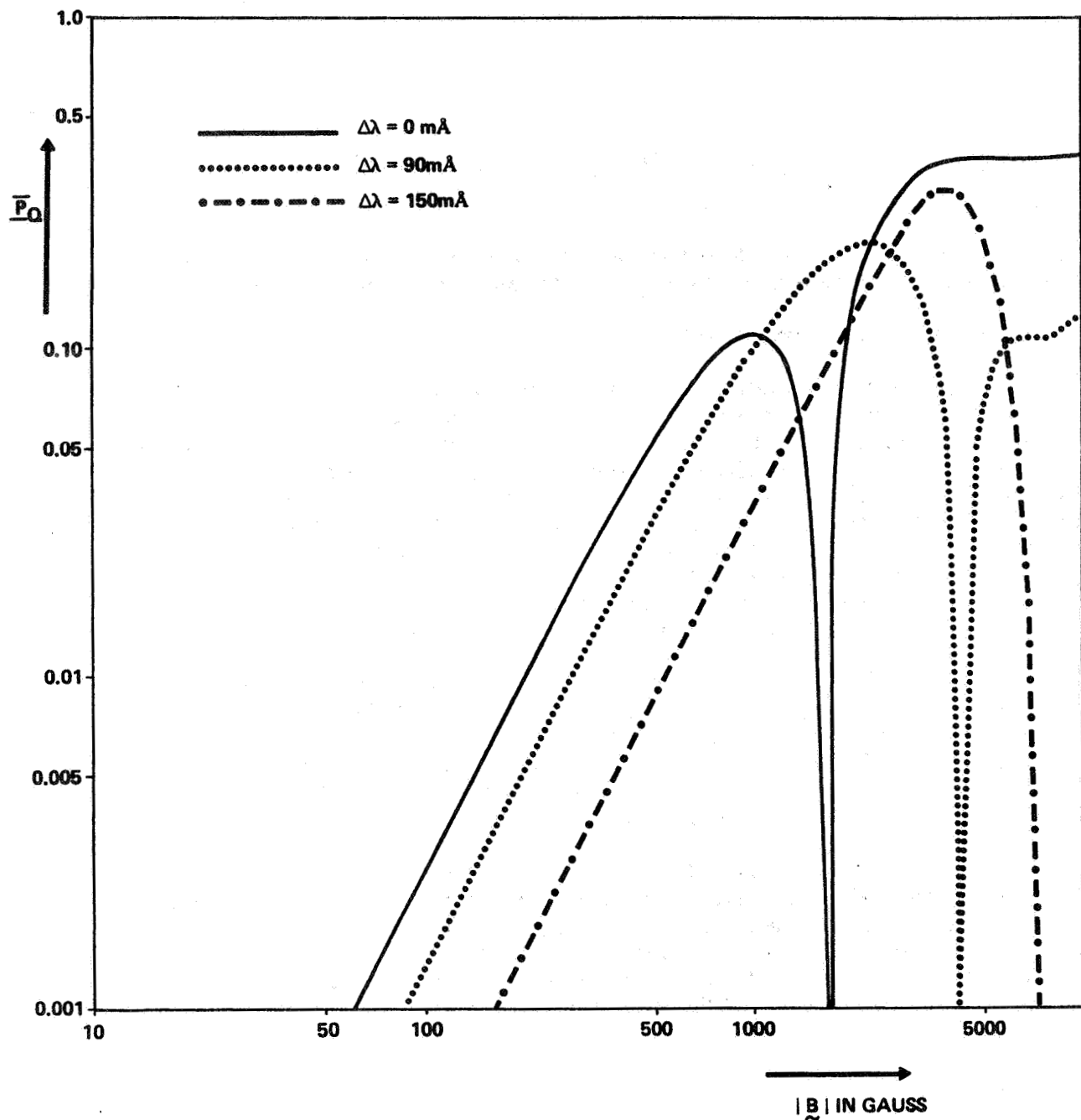


Figure E.4. Extending the regime of the quadratic approximation for the transverse field by tuning the spectral filter. The calculated fractional linear polarization is plotted as a function of field strength for the same parameters of Figure D.3. In this figure, calibration curves are shown for three different positions of the birefringent filter with respect to the center of the spectral line. The solid curve represents the relation between P_0 and $|\tilde{B}|$ for the filter tuned to the center of the line, the dotted curve is for the case of the filter tuned to 90 mA blueward of line center, and the dotted-dashed curve represents the case of the filter tuned to 150 mA in the blue wing. These curves show that the regime where the quadratic approximation relating B_T to the square root of P_0 is extended to higher field strengths as the filter is tuned further into the wing of the spectral line. However, the sensitivity to weak fields is lost in this process.

$$\psi = \arctan [(C_2 P_Q^{1/2}) / (C_1 P_V)] , \quad (9)$$

can be calculated to a good approximation without regard to how the solar atmospheric structure varies over the field of view.

4. Absolute Calibration

In principal, the quantitative interpretation of a magnetogram can be derived from the theoretical calibration curves discussed in the previous section. In fact, measured values of P_V, P_Q are always less than the values predicted for a given field strength, a factor which results partially from loss of contrast due to scattered light effects. Consequently, absolute calibrations must be done to determine a multiplicative factor "k" to correct all measurements of P_V, P_Q . An absolute calibration can be done by using field strengths determined independently for given areas in the field of view.

In addition to this calibration of the magnitudes of P_V and P_Q , measured values of P_Q must be corrected for a dc noise level in the linear polarization data. A correction for P_V is not so critical: a 10^{-3} noise level in P_V translates into a line-of-sight field equal to ~ 5 G, while this noise in P_Q corresponds to a transverse field of ~ 100 G. This dc noise in P_Q primarily will affect calculations of the field inclination ψ ; without removing its effects there would always be a non-zero transverse component in Equation (9). In an effort to correct for this effect, the dc level noise is estimated by averaging the measured linear polarization P_Q in a non-magnetic region in the magnetogram to obtain the dc noise P_Q^0 . Then the fields B_L and B_T are calculated from the following relations:

$$B_L = C_1 k P_V ,$$

and

$$B_T = C_2 k^{1/2} (P_Q^{1/2} - P_Q^0^{1/2}) .$$

Table E.2. Calibration constants for penumbral model.

$\Delta\lambda$ (mÅ)	C_1	Range (gauss)	C_2	Range (gauss)
0	-	-	2230	0-700
-60	3180	0-900	2590	0-1000
-90	4310	0-1500	3190	0-1300
-120	6690	0-2600	4240	0-1800
-150	-	-	5640	0-2800

Table E.3. C_2/C_1 for different atmospheric models.

$\Delta\lambda$ (mÅ)	Photosphere	Plage	Penumbra	Umbra
-60	0.804	0.813	0.814	0.949
-90	0.699	0.662	0.740	0.911
-120	0.554	0.496	0.634	0.829

REFERENCES

- Avery, L. W. and House, L. L.: 1968, Astrophys. J. **152**, 493.
- Bray, R. J., and Loughhead, R. E.: 1964, Sunspots, John Wiley and Sons, New York.
- Beckers, J. M.: 1969, Solar Phys. **9**, 372.
- Hagyard, M. J.: 1970, NASA TM X-64541.
- Hagyard, M. J.: 1971, Solar Phys. **16**, 286.
- Hagyard, M. J. and Teuber, D.: 1978, Solar Phys. **57**, 267.
- Harvey, J. and Livingston, W.: 1969, Solar Phys. **10**, 283.
- Holweger, H.: 1967, Z. Astrophys. **65**, 365.
- Kjeldseth-Moe, O.: 1968, Solar Phys. **4**, 267.
- Kjeldseth-Moe, O. and Maltby, P.: 1969, Solar Phys. **8**, 275.
- Landi Degl'Innocenti, E.: 1976, Astron. Astrophys. Suppl. **25**, 379.
- Mihalas, D.: 1970, Stellar Atmospheres, W. H. Freeman, San Francisco, p. 16.
- Pecker, J. C.: 1965, in C. De Jager (ed.), The Solar Spectrum, D. Reidel Publishing Company, p. 80.
- Shurcliff, W. A.: 1962, Polarized Light, Harvard University Press, Cambridge, Massachusetts.
- Stenflo, J. O.: 1975, Solar Phys. **42**, 79.
- Unno, W.: 1956, Pub. Astron. Soc. Japan **8**, 108.

APPENDIX F

Measurement of Doppler Velocities with the SAMEX Magnetograph

1. Introduction

Observations of velocities at the solar surface provide a foundation for the study of dynamic processes in the Sun. The general flow patterns are complex and include the Evershed flow in sunspots ($0.5-1.0 \text{ km s}^{-1}$) and the downflows and horizontal motions in supergranules ($0.1-0.4 \text{ km s}^{-1}$). Recent observations obtained with the SOUP instruments on Spacelab 2 (Title et al., 1986) indicate the horizontal motions of granules are on the order of $1-2 \text{ km s}^{-1}$ and outflows from sunspots are in the range $0.2-0.5 \text{ km s}^{-1}$.

The photospheric velocity field must certainly play a significant role in the evolution of active region magnetic fields and their buildup into energetically stressed configurations. As the photospheric flows move the footpoints of the magnetic field which is frozen into the solar plasma, the magnetic field must evolve and adjust to these motions. Thus, an important aspect in the study of solar activity is this response of the magnetic field to the photospheric flows and the non-potential configurations of the field that are a direct result of these motions.

The SAMEX magnetograph will contribute significantly to our understanding of this dynamical coupling by measuring both the changing magnetic field at the photosphere and the driving motions that contribute to these changes leading up to the onset of flares. The SAMEX instrument will measure Doppler velocities in the photosphere from observations of spectral intensities in the red and blue wings of selected spectral lines. In this Appendix we outline the analysis necessary to derive line-of-sight velocities from these observed intensities and discuss methods that can be used to calibrate the Dopplergrams.

2. Method for the Measurement of Doppler Velocities

The procedure to be described is applicable to the analysis of spectral intensities measured over spectral passbands larger than the half-width of the spectral line. In this case only three positions of the passband with respect to the spectral line are necessary to derive the Doppler shift. This simplified Doppler analysis is thus well suited to a spectral filter with a 120 mÅ bandpass.

The analysis we use follows the procedure of Henze (1984), but it is modified to consider measurements of intensities from an absorption line instead of an emission feature, and from observations using a filter instead of a spectrograph. The lines are assumed to have normalized gaussian profiles with a thermal

Doppler width of $\Delta\lambda_D$ and a wavelength offset $\Delta\lambda_V$ from the central wavelength of the line due to the Doppler shift. The line intensity profile for absorption per unit wavelength as a function of wavelength $I_a(\lambda)$ is measured relative to the continuum (see Figure F.1) and is given by the expression

$$I_a(\lambda) = \frac{I_o}{\Delta\lambda_D \sqrt{\pi}} e^{-\left(\frac{\lambda - \Delta\lambda_V}{\Delta\lambda_D}\right)^2}.$$

The usual gaussian standard deviation σ is related to the thermal Doppler broadening $\Delta\lambda_D$ by

$$\sigma = \Delta\lambda_D / 2.$$

The intensity $I(\lambda)$ per unit wavelength that is actually measured is $I_c(\lambda) - I_a(\lambda)$, where $I_c(\lambda)$ is the intensity of the continuum per unit wavelength. The measured signal is the unabsorbed intensity integrated over the bandpass of the filter, so we will derive the Doppler shift $\Delta\lambda_V$ in terms of the unabsorbed intensity $I(\lambda)$ rather than the absorbed intensity $I_a(\lambda)$.

The line profiles for the Fe I lines at 5250.2 and 5250.6 Å are shown in Figure F.1. A 120 mÅ bandpass filter is shown at three positions about the 5250.6 Å line. The measured intensities are I_c , I_b , and I_r . The measurement of I_c is obtained with the filter positioned in the local continuum near the two absorption lines; in the nomenclature defined above, I_c is given by the integral of $I_c(\lambda)$ over the bandpass of the filter. The intensity measurements I_r and I_b are taken in the red (r) and blue (b) wings of the spectral line with the filter position in the two wings separated by the width of the filter passband and centered about the unshifted wavelength of the line (i.e., the wavelength for zero Doppler velocity). For the present discussion, we will assume the passband of the spectral filter is wide enough to use the approximation of a semi-infinite square passband. Following the analysis of Henze, let the Doppler shift signal R_v be defined as

$$R_v = (I_r - I_b) / (I_r + I_b - 2 I_c).$$

R_v is thus determined entirely in terms of directly measured quantities. We will now show, for the case of a gaussian line profile, that the Doppler wavelength shift, $\Delta\lambda_V$, is given in terms of the thermal Doppler width of the line, $\Delta\lambda_D$, and the inverse error function of R_v , $\text{Inverf}(R_v)$.

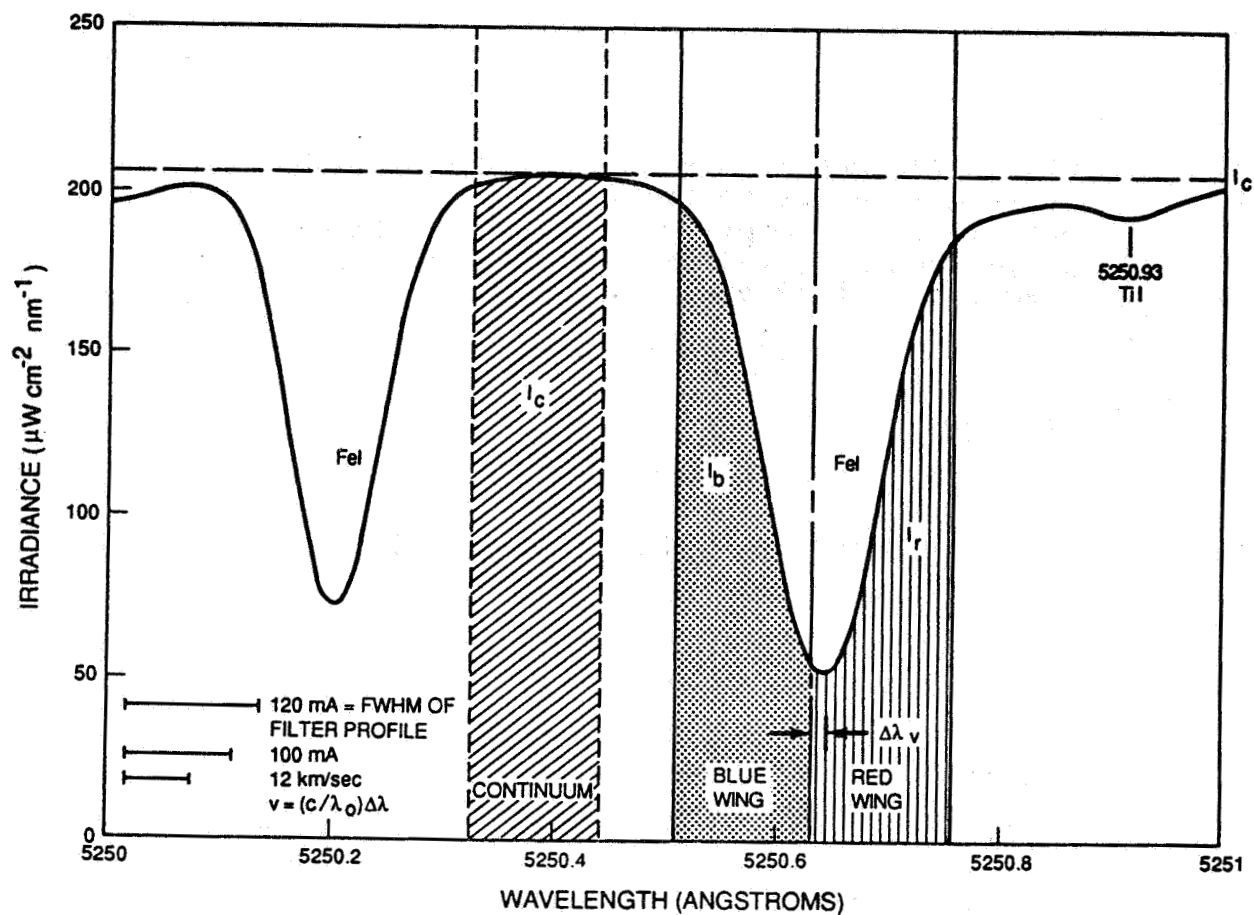


Figure F.1. Spectral profiles of the Fe I 5250.2 and 5250.6 solar absorption lines. The solar irradiance is plotted versus wavelength in the spectral interval between 5250 and 5251 Å. The intensities I_c , I_b , and I_r (identified in the text) are measured with the SAMEX magnetograph to derive the Doppler velocities over the field of view.

The measured intensity with the slit in the blue wing of the absorption line, I_b , is given by the intensity of $I(\lambda)$ integrated over the width s of the slit:

$$I_b = \int_{-s}^0 I(\lambda) d\lambda = \int_{-s}^0 I_c(\lambda) d\lambda - \int_{-s}^0 I_a(\lambda) d\lambda .$$

By definition, the integral of $I_c(\lambda)$ over s is just I_c . Then, since the absorption profile goes rapidly to zero for large wavelengths, i.e., for all λ blueward of the slit, we can approximate the last integral in the above equation by an integral over all negative (blue) wavelengths, so that we can write

$$I_b = I_c - \int_{-\infty}^0 I_a(\lambda) d\lambda .$$

A similar expression can be developed for the measured intensity with the slit in the red wing of the line:

$$I_r = I_c - \int_0^{+\infty} I_a(\lambda) d\lambda .$$

The difference in the measured intensities, $I_b - I_r$, is thus given by

$$I_b - I_r = \frac{I_0}{\Delta\lambda_D \sqrt{\pi}} \left[- \int_{-\infty}^0 e^{-\left(\frac{\lambda - \Delta\lambda_V}{\Delta\lambda_D}\right)^2} d\lambda + \int_0^{+\infty} e^{-\left(\frac{\lambda - \Delta\lambda_V}{\Delta\lambda_D}\right)^2} d\lambda \right] .$$

Let

$$t = \frac{\lambda - \Delta\lambda_V}{\Delta\lambda_D} .$$

With this substitution, we have

$$I_b - I_r = \frac{I_0}{\sqrt{\pi}} \left[- \int_{-\infty}^{-z} e^{-t^2} dt + \int_{-z}^{+\infty} e^{-t^2} dt \right] ,$$

where

$$z = \Delta\lambda_V / \Delta\lambda_D .$$

In terms of the error function $\text{Erf}(y)$, where

$$\text{Erf}(y) = \frac{2}{\sqrt{\pi}} \int_0^y e^{-t^2} dt ,$$

we can now write

$$I_b - I_r = I_o \text{ Erf}(z) .$$

Now I_o is just the normalization factor for the absorption profile, that is,

$$I_o = \int_{-\infty}^{+\infty} I_a(\lambda) d\lambda \approx \int_{-s}^0 [I_c(\lambda) - I(\lambda)] d\lambda + \int_0^{+s} [I_c(\lambda) - I(\lambda)] d\lambda ,$$

or

$$I_o = 2 I_c - (I_r + I_b) .$$

Finally, then, we have

$$Rv = (I_r - I_b)/(I_r + I_b - 2 I_c) = \text{Erf}(\Delta\lambda_V/\Delta\lambda_D) .$$

In terms of the inverse error function, Inverf,

$$\Delta\lambda_V = \Delta\lambda_D \text{ Inverf}(Rv) ,$$

which is the desired relation for the Doppler shift in terms of the Doppler shift signal Rv .

Now the series expansion for the error function is

$$\text{Erf}(y) = \frac{2}{\sqrt{\pi}} \sum_{n=0}^{\infty} \frac{(-1)^n y^{2n+1}}{n! (2n+1)} .$$

Hence, for $y \ll 1$,

$$\text{Erf}(y) \approx 2y/\sqrt{\pi} .$$

Thus, if the Doppler shift is much smaller than the thermal broadening of the line profile, then $\Delta\lambda_V/\Delta\lambda_D \ll 1$, and we can write

$$\Delta\lambda_V \approx \frac{\sqrt{\pi}}{2} \Delta\lambda_D Rv .$$

3. Measurement of the Doppler Width

The thermal Doppler width of the line can also be determined with two additional measurements by using the derived relation

$$\text{Inverf}(Rv) = \Delta\lambda_V / \Delta\lambda_D$$

with data at two different settings of the paired filter measurements (four filter positions). One thus obtains two equations which will allow the two unknowns $\Delta\lambda_V$ and $\Delta\lambda_D$ to be calculated:

$$\text{Inverf}(Rv^{(1)}) = \Delta\lambda_V^{(1)} / \Delta\lambda_D^{(1)} ,$$

and

$$\text{Inverf}(Rv^{(2)}) = \Delta\lambda_V^{(2)} / \Delta\lambda_D^{(2)} .$$

If we assume the thermal Doppler width is constant ($\Delta\lambda_D^{(1)} = \Delta\lambda_D^{(2)}$), we obtain an equation for the unknown Doppler width in terms of measured quantities by subtracting the above equations:

$$\Delta\lambda_D = (\Delta\lambda_V^{(1)} - \Delta\lambda_V^{(2)}) / [\text{Inverf}(Rv^{(1)}) - \text{Inverf}(Rv^{(2)})] .$$

Since $(\Delta\lambda_V^{(1)} - \Delta\lambda_V^{(2)})$ is just the wavelength shift of the filter between the two sets of observations and thus a known quantity, the Doppler width of the line is determined. Therefore, measurements of intensities at five settings of the spectral filter can determine the line-of-sight velocity. If the thermal width of the line is assumed known, then only three measurements are needed.

A second method of calibration for the Doppler shift uses the time-varying component of the spacecraft orbital velocity V_{sc} along the line-of-sight from the spacecraft to the Sun (Henze, 1984). This method is applicable if the wavelength shift caused by the orbital velocity V_{sc} is a significant fraction of the spectral line width. For a near-Earth orbit at 500 km, V_{sc} has a maximum value of 7.62 km s^{-1} ; this velocity gives a wavelength shift of 133 mÅ for the Fe I 5250.6 Å line. Since the equivalent width of this line is 104 mÅ, the line is well suited to this calibration method. For comparison, the Zeeman splitting for a line formed in a magnetic field of intensity B is given by

$$\Delta\lambda_B = (4.67 \times 10^{-13} \text{ G}^{-1} \text{ \AA}^{-1}) g B \lambda^2 ,$$

where g is the Landé factor. For the Fe I 5250.6 \AA line with $g = 1.5$,

$$\Delta\lambda_B = (1.93 \times 10^{-5} \text{ G}^{-1} \text{ \AA}) B .$$

Thus, for a field strength of 500 G, $\Delta\lambda_B = 9.6 \text{ m\AA}$, which is a small fraction, ~ 0.09 , of the equivalent width.

Two other conditions must hold for this method to work: the Doppler width $\Delta\lambda_D$ must remain constant over the interval that the measurements are taken and there should be no large secular or long-duration changes in the line-of-sight velocity of the solar region under observation during this time interval. The method then calls for measurements of the Doppler signal Rv at a number of times during one complete spacecraft orbit. At each point in time, the measured Doppler signal can be related to the Doppler shift of the line:

$$\Delta\lambda_V(t) = \Delta\lambda_O + \lambda_O v_s/c + \lambda_O V_{sc}(t)/c = \Delta\lambda_D \text{ Inverf}[Rv(t)] ,$$

where $\Delta\lambda_O$ represents any (constant) offset of the pair of slits with respect to the center of the spectral line (λ_O) and v_s is the line-of-sight velocity of the observed solar area. At each point in time that a measurement is made, the spacecraft velocity is known (or calculated). If v_s is constant, we can write

$$\text{Inverf}[Rv(t)] = a_0 + a_1 V_{sc}(t) .$$

If $Rv(t)$ is measured at many different times over a large range of the spacecraft velocity, then a least-square fitting procedure will yield the parameters a_0 and a_1 , where

$$a_1 = \lambda_O / (c \Delta\lambda_D)$$

and

$$a_0/a_1 = v_s + c \Delta\lambda_O / \lambda_O .$$

With the parameters a_0 and a_1 determined from the observations, the time-averaged Doppler width of the line can be derived at each pixel location.

4. Derivation of Uncertainties

To derive the uncertainty in the Doppler shift, $\delta\Delta\lambda_V$, we use the approximation

$$\Delta\lambda_V \approx \frac{\sqrt{\pi}}{2} \Delta\lambda_D Rv$$

and assume the uncertainties in $\Delta\lambda_D$ and Rv are independent. Then we have the following relation for the fractional uncertainty in Doppler shift:

$$\frac{\delta\Delta\lambda_V}{\Delta\lambda_V} = \left[\left(\frac{\delta\Delta\lambda_D}{\Delta\lambda_D} \right)^2 + \left(\frac{\delta Rv}{Rv} \right)^2 \right]^{1/2}.$$

We further assume that the dominant uncertainty in calculating $\delta\Delta\lambda_V$ is in the measurement of Rv rather than in $\Delta\lambda_D$, so that we have

$$\frac{\delta\Delta\lambda_V}{\Delta\lambda_V} \approx \frac{\delta Rv}{Rv}.$$

The uncertainty in Rv is given by

$$\delta Rv = \left[\left(\frac{dRv}{dI_b} \delta I_b \right)^2 + \left(\frac{dRv}{dI_r} \delta I_r \right)^2 \right]^{1/2}.$$

For

$$Rv = (I_r - I_b) / (I_r + I_b - 2I_c),$$

we have

$$\delta Rv = 2 \left[\frac{I_c - I_r}{I_o^2} \cdot \delta I_b^2 + \frac{I_c - I_b}{I_o^2} \cdot \delta I_r^2 \right]^{1/2}.$$

If we assume

$$\delta I_b \approx \delta I_r \equiv \delta I,$$

then

$$\delta Rv = \frac{2 \delta I}{I_o^2} \left[(I_c - I_r)^2 + (I_c - I_b)^2 \right]^{1/2}.$$

This can be rewritten as

$$\delta R_V = \frac{2 \delta I}{I_0} \left[1 - \frac{2(I_c - I_r)(I_c - I_b)}{[(I_c - I_r) + (I_c - I_b)]^2} \right]^{1/2}.$$

Assuming

$$I_c - I_r \approx I_c - I_b,$$

this reduces to

$$\delta R_V \approx \sqrt{2} \delta I / I_0.$$

Substituting this into the equation for the fractional uncertainty in $\Delta\lambda_V$, we have

$$\frac{\delta \Delta\lambda_V}{\Delta\lambda_V} \approx \frac{\sqrt{2} \delta I}{R_V I_0}.$$

Using the approximate expression for $\Delta\lambda_V$,

$$\Delta\lambda_V \approx \frac{\sqrt{\pi}}{2} \Delta\lambda_D R_V,$$

we can write

$$\delta \Delta\lambda_V \approx (\pi/2)^{1/2} \frac{\delta I}{I_0} \Delta\lambda_D.$$

If the width of the slit (or the FWHM of the filter) is approximately equal to the equivalent width of the line, then

$$I_0 \approx 2I_r \approx 2I_b \equiv 2I,$$

and we then have

$$\delta \Delta\lambda_V \approx \frac{\sqrt{2\pi}}{4} \frac{\delta I}{I} \Delta\lambda_D.$$

In the actual process of obtaining the measurements, the detector system is the source of the uncertainty δI in the measurement of I . For a photon-counting detector, the counting statistics obey a Poisson distribution. Thus,

$$\frac{\delta I}{I} = (S/N)^{-1}$$

where S/N is the signal-to-noise ratio of the photon-counting detector. Therefore,

$$\delta \Delta \lambda_V \approx \frac{\sqrt{2\pi}}{4} \Delta \lambda_D / (S/N).$$

Finally, using the relation between the FWHM and Doppler width $\Delta \lambda_D$ for a gaussian profile,

$$\text{FWHM} = 0.83 \Delta \lambda_D ,$$

we have

$$\delta \Delta \lambda_V \approx 0.75 \text{ FWHM} / (S/N) .$$

Taking the FWHM of 120 mÅ for the Fe I 5250.6 Å line and assuming a S/N of 10^3 , the uncertainty in the Doppler shift is approximately 0.1 mÅ. This means there is an uncertainty of 6 m s⁻¹ in the measurement of the Doppler velocity. This uncertainty is well below the uncertainty encountered in tuning the spectral filter, ~1 mÅ. Thus, the desired accuracy of 60 m s⁻¹ should not be compromised by this method of measuring Doppler velocities.

More realistic line profiles can be used in the Doppler analysis, and corrections included that account for the presence of other spectral lines, e.g., the Ti I line at 5250.93 Å. In this more detailed analysis, a set of calibration curves would result, but the derivation of the Doppler measurements would follow the general methodology of the development outlined above.

REFERENCES

Henze, W., Jr.: 1984, Research in Solar Physics: Some Techniques for Analyzing Data from the Ultraviolet Spectrometer and Polarimeter, SP84-MSFC-2726, Teledyne Brown Engineering, Huntsville, AL, p. 15.

Title, A. M., Tarbell, T. D., Simon, G. W., and the SOUP Team: 1986, Adv. Space Res. 6(8), 253.

★U.S. GOVERNMENT PRINTING OFFICE: 1988- 526-064/86005

TECHNICAL REPORT STANDARD TITLE PAGE

1. REPORT NO. NASA TM-4048		2. GOVERNMENT ACCESSION NO.		3. RECIPIENT'S CATALOG NO.	
4. TITLE AND SUBTITLE The SAMEX Vector Magnetograph - A Design Study for a Space-Based Solar Vector Magnetograph				5. REPORT DATE June 1988	
				6. PERFORMING ORGANIZATION CODE ES52	
7. AUTHOR(S) M. J. Hagyard, G. A. Gary, and E. A. West				8. PERFORMING ORGANIZATION REPORT #	
9. PERFORMING ORGANIZATION NAME AND ADDRESS George C. Marshall Space Flight Center Marshall Space Flight Center, AL 35812				10. WORK UNIT, NO. M-589	
				11. CONTRACT OR GRANT NO.	
12. SPONSORING AGENCY NAME AND ADDRESS National Aeronautics and Space Administration Washington, DC 20546				13. TYPE OF REPORT & PERIOD COVERED Technical Memorandum	
				14. SPONSORING AGENCY CODE	
15. SUPPLEMENTARY NOTES Prepared by Space Science Laboratory, Science & Engineering Directorate					
16. ABSTRACT This report presents the results of a pre-phase A study performed by the Marshall Space Flight Center's (MSFC) Solar Science Branch for the Air Force Geophysics Laboratory (AFGL) to develop a design concept for a space-based solar vector magnetograph and hydrogen-alpha telescope. As two of the three core instruments for the proposed AFGL Solar Activity Measurements Experiments (SAMEX) satellite, these instruments were designed to provide high-resolution observations of the solar magnetic field in the photosphere and chromosphere. The MSFC ground-based vector magnetograph served as the prototype for the space-based instrument. The primary scientific objective of the SAMEX mission is to understand the role of the Sun's magnetic field in the physics of solar flares. This and other related objectives developed for the SAMEX mission imposed the following instrumental requirements on the vector magnetograph: (1) a temporal resolution of <5 min, (2) a field of view of 4.3 x 8.5 arc min to cover most active regions, (3) a spatial resolution of 0.5 arc sec, (4) a spectral range covering the wavelengths from 524.3 to 525.4 nm, and (5) a polarimetric sensitivity of 10^{-4} to measure the solar magnetic field with greater accuracy than has been done with any other system. The polarimetric accuracy of 10^{-4} was obtained through the use of specially designed optical coatings on the fore-optics and the design of a unique polarimeter. To obtain observations of entire active regions with 0.5 arc sec spatial resolution in less than 5 min required the design of a large-array, charge coupled device (CCD) detector with parallel processing electronics. This report discusses the SAMEX scientific objectives and presents the designs of the optics, polarimeter, spectral filters, and detector that are required to meet these objectives. The report also discusses the numerous trade-offs among spatial resolution, areal coverage, temporal resolution, spectral response, and telemetry that must be considered in achieving the design goals. Spacecraft and mission requirements and requirements for data reduction and analysis are also covered in the report.					
17. KEY WORDS Sun - Space-Based Vector Magnetograph, Solar Activity, Magnetic Fields, Solar Flares; Optics - Polarimetry, Birefringent Filters, Fabry-Perot Etalons; Physics - Zeeman Effect, Charge-Coupled Devices			18. DISTRIBUTION STATEMENT Unclassified--Unlimited Subject Category 92		
19. SECURITY CLASSIF. (of this report) Unclassified		20. SECURITY CLASSIF. (of this page) Unclassified		21. NO. OF PAGES 438	
				22. PRICE A19	